POPULATION GENETICS AND GENOMICS OF EUSOCIAL ANIMALS

A Dissertation
Presented to
The Academic Faculty

by

Linh M Chau

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Biology in the
School of Biological Sciences

Georgia Institute of Technology
December 2017

POPULATION GENETICS AND GENOMICS OF EUSOCIAL ANIMALS

Approved by:

Dr. Michael A.D. Goodisman, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Soojin Yi
School of Biological Sciences
Georgia Institute of Technology

Dr. Patrick McGrath
School of Biological Sciences
Georgia Institute of Technology

Dr. Nicole Gerardo
Department of Biology
Emory University

Dr. Joseph R. Mendelson III
Zoo Atlanta

Date Approved: October 20, 2017

ACKNOWLEDGEMENTS

I would like to thank my parents, Mui and Quang, for making sacrifices and their large emphasis on education throughout my life. My sister, Emma, also.

My advisor, Mike Goodisman was a vital part in my pursuit of a PhD by giving me space to make my own mistakes and develop my own interests. His daily mentoring and advice has shaped me and my aims of what I want to do when I grow up. I had the pleasure of having Soojin Yi, Patrick McGrath, Nicole Gerardo, and Joe Mendelson on my committee. Their scientific insight has taught me the importance of the big picture in science and has pushed me to become a better researcher. I'd also like to thank Joe Lachance and Alison Onstine for their helpful discussion and emotional support.

I must also thank the team at CTL and Patrick Bardill for promoting and fostering my interest in education. Without the opportunities they gave me and their mentoring, I wouldn't be on my current career path.

Students in the Goodisman lab have also provided emotional support and scientific insight throughout my PhD. In particular, I'd like to thank Karl Glastad, Carl Dyson, Laurel Jenkins, and Rachel Kutner. Finally, I am grateful to all my friends in Atlanta, especially Chinar Patil, Jessica Pruett, Kawther Abdilleh, Zach Smith, and Lavanya Rishishwar. Without them, I wouldn't have eaten so much food and I wouldn't be the person I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

N_a,	Number of alleles
N_e,	Effective number of alleles
H_o	Observed heterozygosity
H_e	Expected heterozygosity
A	Allelic richness
N_p	Number of private alleles
Ɵ	Weir and Cockerham's theta
F_{ST}	Fixation index
F_{IS}	Inbreeding coefficient
K	Number of putative clusters
ABC	Approximate Bayesian computation
GDA	Genetic Data Analysis Program
CI	Confidence Interval
NMR	Naked mole rat (*Heterocephalus glaber*)
ZA	Zoo Atlanta
SNZP	Smithsonian National Zoological Park
SDZ	San Diego Zoo
PCR	Polymerase chain reaction
PIC	Phylogenetic independent contrasts
APE	Analysis of Phylogenetics and Evolution
Tau	Level of tissue-biased gene expression
PAML	Phylogenetic Analysis by Maximum Likelihood

RNA	Ribonucleic acid
dN	Nonsynonymous substitution rate
dS	Synonymous substitution rate
FDR	False discovery rate

SUMMARY

Major evolutionary transitions have been responsible for the increase of organismal complexity. The latest transition from solitary life to sociality has led to the development of reproductive division of labor in which individuals are divided into castes each responsible for specific tasks. Reproductive castes are responsible for reproduction, while nonreproductive caste members take part in colony maintenance and brood care. This division of labor represents a challenge to selection and has long been of curiosity to researchers.

This dissertation examines the population genetics and genomics of eusociality in a spectrum of eusocial organisms. I use genetic and genomic techniques to learn more about the factors associated with the evolution of eusociality in eusocial insects and mammals. First, I find that population structure of invasive social insects can be shaped by geography. I also examine the population genetics of naked mole rats, one of the only known eusocial mammals, living in captivity in order to understand how captivity can shape the population structure of a eusocial animal. Finally, I examine how the phenomena of gene duplication, which creates new genetic material in the genome, can affect the evolution of castes in eusocial species.

These studies provide insight on an array of population genetic and genomic questions concerning the evolution of eusociality. Therefore, this research unveils trends associated with the evolution of eusociality across a diverse set of eusocial taxa and furthers our understanding of the rare distribution of this social system across the tree of life.

CHAPTER 1. INTRODUCTION

The evolution of life has been marked by a number of major evolutionary transitions (Szathmary and Smith 1995; West et al. 2015). These include the transition from prokaryotic cells to eukaryotes, asexual to sexual reproduction, and solitary life to eusociality. Eusociality shares common features with the major evolutionary transitions in that individual units that were capable of independent replication join together to form larger units that can only replicate as parts of the larger unit (Batra 1966; Michener 1969; Wilson 1971). Each evolutionary transition has been coupled with a division of labor, which can lead to increased efficiency of the larger unit and the increase of biological complexity. In eusocial societies, some individuals are responsible for reproduction while others are not, which is termed as the reproductive division of labor.

Eusociality is interesting for multiple reasons. First, eusociality is a classic example of phenotypic plasticity, in which organisms have the ability to change physiological or behavioral traits in response to their local environment (West-Eberhard 1989). Second, the reproductive division of labor and altruism proposes a challenge to the theory of natural selection, since reducing one's reproduction seems inefficient for selection (Darwin 1859). Inclusive fitness theory was introduced by Hamilton to address the problem brought up by eusociality. Hamilton's theory highlights the importance of genetic relatedness in the evolution of altruism (Hamilton 1964). Therefore, members of a colony are selected to help related individuals.

The ecological dominance of one group of eusocial organisms, the social insects, has been attributed to eusociality. Task specialization, due to the caste system, promotes

efficiency in a colony, which may aid social organisms to play a diverse set of roles in terrestrial environments (Wilson 1976; Jeanne 1986; Holldobler and Wilson 1990; Mooney et al. 2015). Some social insects can act as soil turners, nutrient distributors, seed dispersers, and they can play a large set of roles in local food webs (Holldobler and Wilson 1990). Social insects like ants, termites, social wasps, and bees also make up a large number of invasive species (Holldobler and Wilson 1990; Bourke AFG 1995; Crozier RH 1996; Tsutsui et al. 2000; Tsutsui and Suarez 2003; Wilson and Holway 2010; Ascunce et al. 2011; Schmid-Hempel et al. 2014; Chau et al. 2015). It is thought that the success of invasive social insects is due to their flexibility arising from having both individual and colony level responses that allow them to react and conquer new environments (Moller 1996). Therefore, eusociality may play a large role in the ecological dominance of eusocial animals.

We can examine the ultimate and proximate causes of the evolution of eusociality. Ultimate causes emphasize the larger, evolutionary drivers of eusociality, whereas proximate causes focus on specific, mechanistic causes. Some of the ultimate causes of eusociality include selection pressures and ecological factors that promote the evolution of group living. Group living can be beneficial because individuals can pool resources and labor, which may increase the likelihood of survival and passage of their genes to the next generation (Faulkes and Bennett 2013). For example, in some species, nests may be hard to construct, making it a limited resource. Staying at the natal nest to inherit this resource may be a better option than finding a new nest, favoring individuals who stay compared to those that disperse. Such competition for scarce resources may limit dispersal and provide incentive for social individuals to remain in their natal colony (Queller and Strassmann

1998). Other selection pressures possibly associated with eusociality include predation and parasitism (Lin and Michener 1972; Wcislo and Cane 1996; Rehan et al. 2011). Solitary species leave nests unguarded to obtain resources, which leaves their brood unprotected from parasites and predators. Therefore, group living allows for guards to stay at the nest and provide protection for vulnerable brood (Gadagkar 1990, 1991). Overall, such ecological constraints can induce shifts in behavior and possibly promote changes in social structure.

Unlike ultimate causes, proximate causes approach the evolution of eusociality at a more mechanistic level, emphasizing the development and genetics of group living. One example of a mechanism leading to the evolution of eusociality is kin recognition. Being able to recognize kin allows individuals to identify who to cooperate with. This can happen through a number of mechanisms, which include olfactory signals and the green beard effect. The social b supergene in fire ants, which signals whether a colony accepts multiple queens, represents a great example of the green beard effect. Also, other mechanisms can predispose species to evolve eusociality. For example, sex determination can alter the relatedness between individuals in a colony. High levels of relatedness may be conducive for the formation of social groups. In the classic example of the honey bee and other haplodiploid social insects, sisters within a colony are 75% related. This high level of relatedness promotes helping between sisters because individuals will gain indirect fitness by helping their highly related sisters. Other proximate causes possibly include parental care and helpers.

There are a number of models that address the origin and evolution of eusociality at the molecular level (Rehan and Toth 2015). A handful of these models focus on changes

in gene expression that promote the evolution of alternative phenotypes. One example is the genetic tool kit hypothesis which predicts that there is a shared toolkit of molecular and physiological processes used in the evolution of the caste system across several, independently evolved social organisms (True and Carroll 2002; Rehan and Toth 2015). Some models include changes in the genome which lead the development of alternative phenotypes. For example, the novel genes hypothesis proposes that novel protein coding genes are co-opted into the evolution of eusociality (Rehan and Toth 2015).

With the recent field of genomics, researchers have become more interested in testing models concerning the evolution of sociality in social insects (Robinson et al. 2005). One of the first feats came with the sequencing of the honeybee (*Apis mellifera*) genome. Later came many other social insect genomes, which allowed for comparison between the genomes to understand the evolutionary trends across the social insects (Bonasio et al. 2010; Nygaard et al. 2011; Smith et al. 2011; Suen et al. 2011; Wurm et al. 2011; Kocher et al. 2013; Oxley et al. 2014; Kapheim et al. 2015; Patalano et al. 2015; Sadd et al. 2015; Standage et al. 2016). In addition to genomes, researchers have also become interested in epigenetics, particularly in DNA methylation, and the role it plays in caste differentiation, (Elango et al. 2009; Foret et al. 2009; Bonasio et al. 2012; Herb et al. 2012; Standage et al. 2016). Early data from honeybees peaked interest in the role of DNA methylation in the regulation of gene expression in the development of queen and worker castes (Kucharski et al. 2008; Lyko and Maleszka 2011). With the accumulation of data in a wider variety of social insects, DNA methylation seems to be more labile and not generally associated with the evolution of castes in all social insects (Standage et al. 2016).

Researchers have also been interested in the differences in gene expression associated with the development of alternative phenotypes (Whitfield et al. 2003; Sumner et al. 2006; Bonasio et al. 2010; Terrapon et al. 2014; Berens et al. 2015; Morandin et al. 2015; Patalano et al. 2015). Variation in gene expression has been linked to variation in phenotypic development; therefore, researchers are interested in the genes differentially expressed across castes and other alternative eusocial phenotypes (Pereboom et al. 2005; Sumner et al. 2006; Hoffman and Goodisman 2007; Ometto et al. 2011; Herb et al. 2012; Ferreira et al. 2013; Feldmeyer et al. 2014; Harrison et al. 2015; Morandin et al. 2015). For the aforementioned, genetic tool kit hypothesis, there has been gene expression studies used to support this hypothesis (Berens et al. 2015). Multiple studies carried out across the hymenoptera suggest the presence of a 'loose toolkit' in which there are key pathways associated with social traits across multiple social species, instead of specific genes (Berens et al. 2015). Also, there have been a number of gene expression studies examining the novel genes hypothesis. Past studies have seen that 'novel genes', those that lack homology to known genes or to be restricted to certain taxa, tend to be highly expressed in worker castes (Johnson and Tsutsui 2011; Ferreira et al. 2013; Harpur et al. 2014; Standage et al. 2016).

Most genetic and genomic research on eusociality has been focused on the social insects. However, eusociality has also appeared in other invertebrates like thrips (Crespi 1992), aphids, ambrosia beetles and sponge-dwelling shrimp (Kent and Simpson 1992; Duffy 1996; Stern and Foster 1996). Eusociality also appeared in mammals, such as the naked mole rat (Jarvis 1981) and Damaraland mole rat (Bennett 1990; Jarvis and Bennett 1993) (Faulkes and Bennett 2013).

This thesis aims to study different genetic aspects of eusociality across multiple eusocial species. Currently, most genetic and genomic studies concerning the evolution of eusociality have focused solely on the Hymenoptera. This happens probably because hymenopteran genetics is well developed and there are major differences in biology across the spectrum of eusocial organisms. However, it is good to keep the spectrum of eusocial animals in mind because we might see different genetic aspects potentially associated with the evolution of eusociality. Also, we may gain insight that may explain the distribution of eusociality across different taxa. Therefore, this thesis will examine the population genetics and genomics of multiple eusocial taxa like wasps, across the bees, and naked mole rats.

Chapter two of this dissertation is focused on the population genetics of the invasive wasp, *Vespula pensylvanica* (Chau et al. 2015). Eusocial insects make up a large number of invasive species, so it's interesting to examine the genetic factors that may be associated with their success. Therefore, we surveyed the invasion of *V. pensylvanica*, a yellowjacket wasp native to North America, in the ocean archipelago of Hawaii. Using microsatellite markers, we measured levels of genetic diversity and compared native and invasive populations to examine the genetic changes in a population that occur during a successful invasion. Overall, we saw a lack of genetic variation in *V. pensylvanica's* native range but there was the presence of high genetic differentiation across its introduced range amongst the Hawaiian Islands.

Chapter three investigated the population genetics of the naked mole rat, a eusocial mammal. This study aimed to examine the population structure and levels of genetic diversity that accompany a eusocial animal living in captivity. It also provided insight into

the sex ratio of a eusocial mammal, which is not particularly well studied compared to eusocial insects. We created a set of microsatellite primers and genotyped individuals from three zoos across the United States. There were modest levels of genetic variations in the zoo populations. Also, we saw the presence of population genetic structure, which may reflect the isolation of captive naked mole rat colonies at different zoos. Overall the results of this study may be useful in maintaining eusocial organisms in captivity.

Finally, chapter four is centered the genomic changes associated with eusociality, specifically the role of gene duplication in the evolution of caste and eusociality (Chau and Goodisman). The aim of this chapter overall was to look at the genomic changes potentially associated with the evolution of eusociality and to examine an aspect of the novel genes hypothesis. Interestingly, we saw a relationship between rate of gene duplication and level of sociality in the Apoidea. Also, we saw a relationship between gene duplication and differential expression of alternative phenotypes, which may suggest that gene duplicates are co-opted in the evolution of castes, sexes, and other alternative phenotypes. These results suggest that gene duplicates may be co-opted in the evolution of alternative phenotypes in the honeybee.

This dissertation aims to examine multiple aspects of eusociality across the spectrum of eusocial taxa. Overall, these chapters examined the population genetics and genomics of eusocial organisms. By studying eusocial species like wasps, naked mole rats, and bees, we're adding to the phylogenetic breadth of our understanding of eusociality. This will further help us understand the genetic factors that led to the evolution of eusociality.

CHAPTER 2. POPULATION GENETIC STRUCTURE OF THE PREDATORY SOCIAL WASP *VESPULA PENSYLVANICA* IN ITS NATIVE AND INVASIVE RANGE[1]

2.1 Abstract

Invasive species cause extensive damage to their introduced ranges. Ocean archipelagos are particularly vulnerable to invasive taxa. In this study, we used polymorphic microsatellite markers to investigate the genetic structure of the social wasp *Vespula pensylvanica* in its native range of North America and its introduced range in the archipelago of Hawaii. Our goal was to gain a better understanding of the invasion dynamics of social species and the processes affecting biological invasions. We found that *V. pensylvanica* showed no significant genetic isolation by distance and little genetic structure over a span of 2000 km in its native range. This result suggests that *V. pensylvanica* can successfully disperse across large distances either through natural- or human-mediated mechanisms. In contrast to the genetic patterns observed in the native range, we found substantial genetic structure in the invasive *V. pensylvanica* range in Hawaii. The strong patterns of genetic differentiation within and between the Hawaiian Islands may reflect the effects of geographic barriers and invasion history on gene flow. We also found some evidence for gene flow between the different islands of Hawaii which was likely mediated through human activity. Overall, this study provides insight on how

[1] Chau, L. M., C. Hanna, L. T. Jenkins, R. E. Kutner, E. A. Burns, C. Kremen, and M. A. D. Goodisman. 2015. Population genetic structure of the predatory, social wasp *Vespula pensylvanica* in its native and invasive range. 5:5573-5587.

geographic barriers, invasion history, and human activity can shape population genetic structure of invasive species.

2.2 Introduction

Invasive species are recognized as one of the top threats to the environment (Sakai et al. 2001; Pejchar and Mooney 2009; Kirk et al. 2013; Simberloff et al. 2013). Introduced species can displace native taxa, alter habitats, act as vectors for foreign diseases, and reduce levels of biodiversity (Sakai et al. 2001; Kenis et al. 2009; Brockerhoff et al. 2010; Beggs et al. 2011). Invasive species are often transported to new locations through human mediated methods (Sakai et al. 2001). Consequently, the growing rate of globalization has increased the risk of nonnative species being introduced to new regions (Pejchar and Mooney 2009).

Many social insects are highly successful invasive species (Moller 1996; Tsutsui et al. 2000; Chapman and Bourke 2001; Tsutsui and Suarez 2003; Beggs et al. 2011; Husseneder et al. 2012; Evans et al. 2013; Kirk et al. 2013; Gotzek et al. 2015). Introductions of invasive termites, ants, social bees, and social wasps have caused substantial damage to local ecosystems and economies (Holway et al. 2002; Suarez and Case 2002; Beggs et al. 2011; Evans et al. 2013). The success of social insects as invasive species is likely associated with their social structure, in addition to other factors like their occupation of broad niches, high dispersal power, and effective predator defense (Moller 1996). These traits allow invasive social insects to work efficiently and rapidly increase in density in new environments, raising their likelihood of invasion success (Moller 1996; Smith et al. 2008).

Vespula wasps are particularly notorious invasive social insects. *Vespula* wasps are native to various regions throughout the northern hemisphere but have been introduced to many locations, such as Australia, South America, Hawaii, and New Zealand (Akre et al. 1981; Brockerhoff et al. 2010; Beggs et al. 2011; Monceau et al. 2014). Introductions of *Vespula* wasps have led to negative consequences for their invasive ecosystems (Matthews et al. 2000; Beggs et al. 2011). For example, *Vespula* species are known to compete with native pollinators and carnivores (Brockerhoff et al. 2010; Elliott et al. 2010; Beggs et al. 2011; Hanna et al. 2014b). This phenomenon has serious costs and has been linked to the population decline of native taxa (Elliott et al. 2010).

The western yellowjacket, *Vespula pensylvanica*, has emerged as one of the most destructive invasive *Vespula* species. *V. pensylvanica* is native to the western parts of North America but was recently introduced to all of the major islands in Hawaii (Nakahara 1980; Akre et al. 1981; Visscher and Vetter 2003). The introduction of *V. pensylvanica* to Hawaii has had serious consequences for native Hawaiian fauna. Since Hawaii has no native social insects, introduced *V. pensylvanica* populations have no direct, native, social insect competitors (Wilson 1996). Thus, the introduction of *V. pensylvanica* into Hawaii has led to the displacement of endemic insects and pollinators, such as the Hawaii picture wing fly and *Hylaeus* bees (Foote and Carson 1995; Wilson and Holway 2010; Hanna et al. 2014b). The ecological effects of *V. pensylvanica* are possibly magnified by the increased population density that stems from perennial nests that are common in Hawaiian populations (Nakahara 1980; Gambino 1991; Hanna et al. 2014a).

The purpose of this study is to use genetic markers to gain a greater understanding of the invasion of Hawaii by *V. pensylvanica*. The historical records of the invasion and

presumed consequences of species invasions allow us to make predictions about the population genetic structure of invasive and native *V. pensylvanica*. For example, we expect that invasive populations will harbor less variation than native populations, as is typical for introduced species (Dlugosch and Parker 2008). In addition, we expect that some introduced populations may show evidence of population bottlenecks, which occur if populations undergo reductions in size during the founding process (Cornuet and Luikart 1996; Luikart et al. 1998).

We also predict that *V. pensylvanica* will display genetic isolation by distance across its native range, given the broad distribution of *V. pensylvanica* across western North America and the presumed limited dispersal ability of *Vespula* queens (Masciocchi and Corley 2013). In contrast, we expect little genetic isolation by distance within islands in Hawaii. Introduced Hawaiian populations are believed to have been recently founded from multiple, distinct introduction events, which would be expected to obscure patterns of genetic isolation by distance (Nakahara 1980).

Finally, we predict differences in genetic relationships between *V. pensylvanica* populations on the western Hawaiian Islands of Kauai, Oahu, and the eastern Hawaiian Islands of Molokai, Lanai, Hawaii, and Maui. Populations on Molokai, Lanai, Hawaii, and Maui were colonized in the late 1970s (Nakahara 1980). These populations were thought to have been established by Christmas trees shipments from Oregon (Nakahara 1980). So we expect that these eastern populations will be closely related to each other. In contrast, *V. pensylvanica* was first noted on Kauai and Oahu in 1919 and 1936, respectively (Nakahara 1980). Thus we predict that the populations on Kauai and Oahu will be less related to each other, and to the populations on the eastern islands.

Overall, the goal of this study is to understand the invasion of *V. pensylvanica* across the Hawaiian Islands. Archipelagoes, like Hawaii, serve as models for investigating the interplay between ecological and evolutionary processes in shaping invasion dynamics because they vary in shape, size, degree of isolation, and age (Drake et al. 2002). We are interested in determining how geographic barriers affect population structure and genetic variation in native and invasive habitats. Ultimately, we hope this study will provide insight into the role of geography and the effects of humans on biological invasions.

2.3 Methods

2.3.1 Sampling Scheme

We collected 1364 *V. pensylvanica* workers from their native range in the western part of the United States and their invasive range in Hawaii in 2008 (Table A 1). Native samples were collected from 170 traps within the states of California, Colorado, Oregon, Wyoming, and New Mexico. Invasive samples were collected from 178 traps from the Hawaiian Islands of Kauai, Oahu, Molokai, Lanai, Maui, and Hawaii. Specimens were collected by deploying 5-15 Seabright Yellow Jacket and Wasp Traps ®, separated by ≥ 325 m, for 24-48 hours. The traps were baited with n-heptyl butyrate emitted from controlled-release dispensers (Landolt et al. 2003). Wasps collected in traps were placed in 95% ethanol for subsequent genetic analysis.

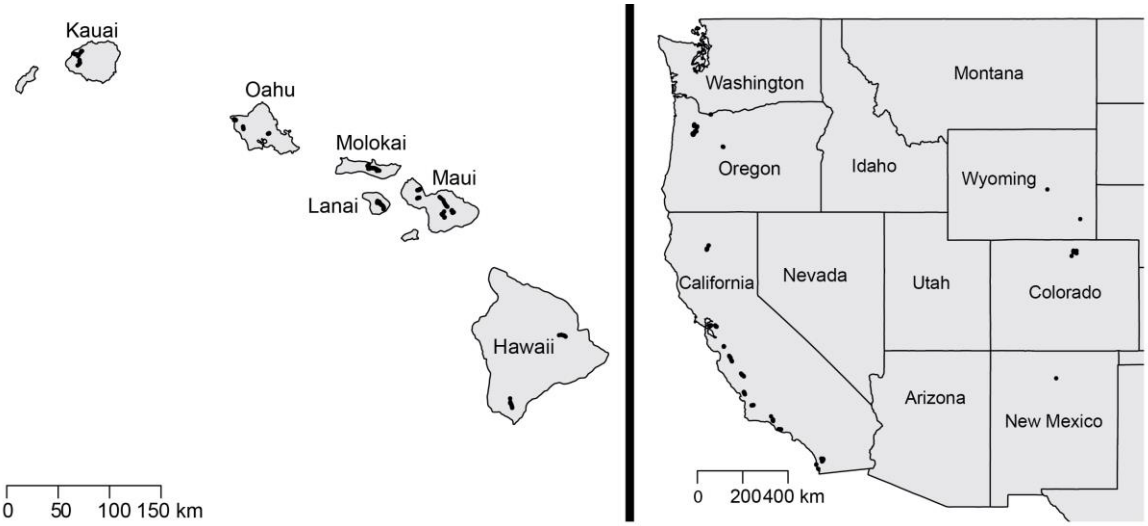

Figure 2-1 Locations of *V. pensylvanica* traps in the invasive (Hawaiian) and native (mainland) range.

Sampling was conducted in a hierarchical manner consisting of four levels: traps, transects, regions, and ranges. Multiple traps were set along more or less linear transects, which spanned up to 14.5 kilometers. Several transects were found within regions, defined as either the focal state in the native habitat or island in the invasive habitat. Regions were then grouped into two distinct ranges; the native range consisted of all states in the mainland of the United States and the invasive range consisted of the Hawaiian Islands (Table 2-1).

Table 2-1 Total numbers of traps and individual V. pensylvanica wasps collected from each transect in the different sampled ranges and regions.

Range	Region	Transect	Number of Traps	Number of Individuals
Native	California	Atascandero	8	38
		Balboa Park	5	21
		BR	8	29
		Corning	3	8
		Diablo	10	43
		La Jolla	2	9
		Lake Shasta	7	27
		Los Padres	9	43
		Morgan Hill	4	20
		Portrero Road	7	23
		Ramona	11	43
		Santa Maria	8	31
		Tilden Park	13	59
		Tres Pinos	8	29
	Colorado	Loveland	2	10
		Outside Fort Collins	4	20
		Outside Larimer County	1	5
		Within Fort Collins	8	35
	New Mexico	Chimayo	1	5
	Oregon	Chemult	1	2
		Columbia River Gorge	3	7
		Klamath Falls	1	4
		Mill City	4	8
		Salem Area	37	131
		Sisters	3	5
	Wyoming	Chugadul Caspar	2	5
Invasive	Hawaii	Kahuku	11	45
		SRA	7	22
	Kauai	Highway-552	11	35
		Makaha Ridge	9	40
	Lanai	Garden of the Gods	1	5
		Monroe Trail	23	101
		Shipwreck	1	1
	Maui	Hosmer Grove	4	4
		Haleakala	7	25
		Maui Iao Valley	10	45
		Olinda Road	13	62
		Waihee Ridge Trail	14	64
		Waipoli Road	9	30
	Molokai	Forest Reserve Road	27	122
		Molakai Kalaupapa 23	2	8
	Oahu	Manana	9	33
		Satellite Road	11	39
		WV	9	23
Total			348	1364

2.3.2 DNA extraction and genotyping

We assayed the multilocus genotype of 1364 *V. pensylvanica* workers at the following 15 microsatellite markers, VMA6, LIST2002, LIST2003, LIST2004, LIST2007, LIST2008, LIST2010, LIST2014, LIST2015, LIST2017, LIST2019, LIST2020, RUFA3, RUFA5, and RUFA19 (Thoren et al. 1995; Daly et al. 2002; Hasegawa and Takahashi 2002). DNA was extracted from the legs of *V. pensylvanica* workers using a modified Chelex protocol (Goodisman et al. 2001). Loci were PCR-amplified with fluorescently-labelled primers (Hoffman et al. 2008). The resulting PCR fragments were subsequently run on an ABI 3100 Genetic Analyzer. Alleles were scored using GeneMarker v 4 (SoftGenetics).

2.3.3 Genetic data analysis

Genetic diversity measures, including number of alleles, effective number of alleles, observed heterozygosity (H_o), and expected heterozygosity (H_e), were calculated with GenAlEx v 6.5 (Peakall and Smouse 2012). We used GENEPOP v 4.3 to test for deviations of genotype frequencies from Hardy-Weinberg equilibrium within each transect (Rousset 2008). The Bonferroni correction was used to adjust for multiple testing.

Our initial analysis detected significant deviations from Hardy-Weinberg equilibrium in 32 transects. We found that most departures were caused by an excess of homozygosity at the locus LIST2002. Microchecker v 2.2 was thus used to detect the presence of null alleles by identifying heterozygote deficiencies at each locus (van

Oosterhout et al. 2006). We confirmed that LIST2002 displayed significant evidence of null alleles in 26 out of 44 transects. Due to the deviation from Hardy-Weinberg equilibrium and the putative presence of null alleles, we eliminated LIST2002 from our analyses. All subsequent statistical tests were thus performed without LIST2002.

We screened for linkage disequilibrium between loci within transects using GENEPOP v 4.3 (Rousset 2008). Default parameters were used for all tests. Allele number and sample size corrected allelic richness were calculated with FSTAT v 2.9.3 (Goudet 1995). Wilcoxon signed-rank tests were used to compare allelic richness between the native and invasive ranges, and also among regions within ranges. We used Friedman tests to compare levels of allelic richness among the Hawaiian Islands in the invasive range.

We estimated Weir and Cockerham's Θ using Genetic Data Analysis (GDA) v 1.1 in order to assess levels of genetic differentiation (Weir and Cockerham 1984; Weir 1996; Holsinger and Weir 2009). Estimates of population structure were obtained at multiple levels including: individuals within traps, traps within transects, transects within regions, and regions within ranges. GDA was also used to calculate 95% confidence intervals (based on 1000 bootstraps) around estimates of Θ. Traps, transects, and regions with less than two samples were excluded from the analyses.

Pairwise measures of F_{st} were calculated between all traps using GENEPOP. These measures of genetic distance were then compared to geographic distance to determine if *V. pensylvanica* displayed evidence for genetic isolation by distance (Wright 1943; Rousset 2008). The significance of the correlation between geographic and genetic distances was

assessed with Mantel tests. These tests were performed with 1000 permutations in the R package vegan v 2.0 (Dixon 2003).

Individuals were assigned to putative populations using Bayesian clustering as implemented by STRUCTURE v 2.2 (Pritchard et al. 2000). To estimate the number of populations (K) present in the native and invasive range, we performed different simulations, each under the assumption of a different K value (1 to 44), representing the total number of transects. For each simulation, we used an admixture model with uncorrelated allele frequencies to account for wasps with mixed ancestry and the LOCPRIOR model to use sampling location to inform clustering. Each simulation was run 10 times with 10,000 steps of burn-in and 50,000 Markov chain Monte Carlo (MCMC) repetitions. The most likely number for K was selected based on log likelihood and the ΔK statistic developed by Evanno et al. (2005) as implemented in STRUCTURE HARVESTER (Pritchard et al. 2000; Evanno et al. 2005; Earl and Vonholdt 2012). For a given set of simulations for each K, CLUMPP v 1.1.2 was used to align the 10 replicate runs (Jakobsson and Rosenberg 2007). Distruct v 1.1 was then used to visualize the results of the clustering process (Rosenberg 2004).

We used the program GeneClass2 to determine the origin of individuals from the invasive range (Piry et al. 2004). We first used assignment tests to determine the likelihood that an individual *V. pensylvanica* wasp in the invasive range was from an identified reference population in native range. For the assignment tests, we used the Bayesian criteria developed by Rannala and Mountain (1997) with an assignment threshold of 0.05. We also used GeneClass2 to exclude native regions as source populations for invasive *V. pensylvanica*. To exclude individuals, we used the Bayesian criteria from Ranala and

Mountain (1997) along with the resampling algorithm from Paetkau et al (2004). We calculated the exclusion probability for each introduced individual with 1000 MCMC simulations and an alpha level of 0.01.

We used the program DIYABC 2.0 (Cornuet et al. 2014) to further understand the invasion process and detect possible source populations of invasive *V. pensylvanica*. DIYABC 2.0 uses approximate Bayesian computation (ABC) which is a Bayesian approach that compares the posterior probabilities of a large number of simulated datasets under given models to those calculated from observed data (Beaumont 2010). For each test, we compared scenarios to find potential source populations of an invasive population and to check for presence of low effective population size after introduction (bottleneck). We compared four scenarios: (1) the focal invasive population was sourced from the native western regions (CA/OR), (2) the focal invasive population was sourced from the native central regions (WY/CO/NM), (3) the focal invasive population was sourced from the native western regions and underwent a bottleneck, and (4) the focal invasive population was sourced from the native central regions and underwent a bottleneck. The models included uniform priors with the following constraints: $t2 > t1$, $db < t2$, and $N1b < N1$. A generalized stepwise mutation model was used for all analyses. Each test generated reference tables with 4 x 105 simulated datasets. Posterior probabilities were estimated for each scenario using polychotomous logistic regression.

Poptree2 was used to generate neighbor-joining (NJ) trees for individuals within transects and regions (Takezaki et al. 2010). Each tree was constructed using Nei's DA distance. Node confidence was assessed using 1000 bootstraps (Nei et al. 1983).

Finally, the program Bottleneck was used to identify populations that may have recently undergone a decrease in population size (Luikart et al. 1998; Piry et al. 1999). This program exploits the principle that allele number is reduced faster than heterozygosity in populations that have recently experienced a reduction in effective population size. We used the Wilcoxon test with the two phase mutation model (TPM), which is recommended for microsatellite datasets with small sample sizes per population and low numbers of polymorphic loci, to determine if populations showed significant evidence of having passed through a recent bottleneck.

2.4 Results

2.4.1 Genetic diversity

We investigated if levels of genetic diversity differed between native and invasive *V. pensylvanica*. We found that wasps from the native range had significantly higher allelic richness (Wilcoxon Sign-Rank Test; $p < 0.001$) and effective number of alleles ($p = 0.001$) than wasps from the invasive range (Table 2-2). We also found that number of private alleles in the native range (30 total) differed significantly ($p = 0.0219$) from the number of private alleles in the invasive range (8 total). Finally, the native range had a significantly higher level of expected heterozygosity ($p = 0.001$) and observed heterozygosity ($p = 0.023$) when compared to the invasive range. Overall, these results suggest that there is a slightly, but significantly, higher level of genetic diversity in the native range than the invasive range of *V. pensylvanica*.

Table 2-2 Measures of genetic diversity at microsatellite loci for native and invasive *V. pensylvanica*, including number of alleles (N_a), effective number of alleles (N_e), observed heterozygosity (H_o), expected heterozygosity (H_e), allelic richness (A), and number of private alleles (N_p).

Locus		N_a	N_e	H_o	H_e	A	N_p
LIST2003	Native	17	3.815	0.719	0.738	16.515	7
	Invasive	11	3.293	0.615	0.696	10.773	1
LIST2004	Native	10	6.081	0.848	0.836	9.996	2
	Invasive	8	5.268	0.721	0.810	8.000	0
LIST2007	Native	19	8.993	0.857	0.889	18.656	6
	Invasive	13	7.195	0.701	0.861	12.800	0
LIST2008	Native	11	3.619	0.704	0.724	10.807	5
	Invasive	7	3.766	0.658	0.734	6.798	0
LIST2010	Native	17	8.449	0.852	0.882	16.611	7
	Invasive	10	6.005	0.694	0.833	10.000	0
LIST2014	Native	26	4.348	0.701	0.770	25.735	11
	Invasive	15	4.841	0.697	0.793	14.954	0
LIST2015	Native	9	4.380	0.726	0.772	8.971	1
	Invasive	9	3.766	0.605	0.734	8.576	1
LIST2017	Native	9	1.681	0.382	0.405	8.631	3
	Invasive	6	1.827	0.415	0.453	5.792	0
LIST2019	Native	6	1.671	0.390	0.402	6.000	1
	Invasive	7	2.101	0.521	0.524	6.588	2
LIST2020	Native	25	10.593	0.907	0.906	24.301	10
	Invasive	15	6.518	0.770	0.847	14.911	0
RUFA19	Native	16	5.726	0.833	0.825	15.997	2
	Invasive	14	3.833	0.674	0.739	13.752	0
RUFA3	Native	30	7.757	0.632	0.871	30.000	13
	Invasive	21	5.032	0.546	0.801	20.683	4
RUFA5	Native	18	6.472	0.821	0.846	17.938	9
	Invasive	10	5.289	0.632	0.811	9.446	1
VMA6	Native	28	12.892	0.889	0.922	27.764	7
	Invasive	22	7.528	0.748	0.867	21.453	1
Mean for all loci	Native	17.214	6.177	0.733	0.771	16.994	6.000
	Invasive	12.000	4.733	0.643	0.750	11.752	0.714

We next investigated differences in genetic diversity between different islands in the invasive Hawaiian range (Table A 2). There were significant differences in observed

heterozygosity (Friedman test; $p = 0.0023$), expected heterozygosity ($p < 0.001$), effective number of alleles ($p < 0.001$), and number of private alleles ($p < 0.001$) among the islands. Interestingly, the island of Hawaii had the highest effective number of alleles, observed heterozygosity, expected heterozygosity, and allelic richness. Maui had highest number of private alleles. In contrast, Kauai had the lowest observed heterozygosity, expected heterozygosity, allelic richness, effective number of alleles, and number of private alleles.

Recently bottlenecked populations may display an excess of heterozygosity compared to expected heterozygosity calculated from observed allele number (Cornuet and Luikart 1996; Luikart et al. 1998). We found significant excesses of heterozygosity in Maui ($p = 0.034$) and New Mexico ($p < 0.001$). Additionally, there were marginally significant excesses of heterozygosity present in the islands of Hawaii ($p = 0.052$), Lanai ($p = 0.052$), and Molokai ($p = 0.086$). In contrast, Oahu ($p = 0.852$) and Kauai ($p = 0.380$) displayed no signs of bottlenecks. Overall, there is some evidence for population bottlenecks in the eastern islands of Hawaii but not the western islands

2.4.2 Genetic differentiation

We measured genetic differentiation among hierarchically structured traps, transects, regions, and ranges of *V. pensylvanica*. We first considered measures of genetic structure for all individuals, combining data from both the native and invasive ranges. We found that f, which measures true inbreeding within populations, was relatively low, albeit significant (Table 2-3). We also uncovered significant differentiation at higher levels of sampling structure. Differentiation between transects within regions, as well as regions

within ranges, was moderate. In spite of these results, we found no significant genetic differentiation between the native and invasive ranges (Table 2-3).

Table 2-3 F-statistics (and 95% confidence intervals) for different levels of genetic structure in V. pensylvanica.

	All Samples	Native Range	Invasive Range
F	0.029 (0.004 – 0.07)	0.012 (−0.020 – 0.065)	0.001 (−0.018 – 0.041)
F	0.112 (0.082 – 0.156)	0.049 (0.017 – 0.098)	0.167 (0.139 – 0.224)
Θ_{traps}	-	0.037 (0.031 – 0.043)	0.167 (0.147 – 0.196)
$\Theta_{transects}$	0.085 (0.075 – 0.098)	0.011 (0.009 – 0.013)	0.157 (0.133 – 0.184)
$\Theta_{regions}$	0.036 (0.030 – 0.042)	0.003 (0.001 – 0.005)	0.137 (0.111 – 0.164)
Θ_{ranges}	−0.003 (−0.008 – 0.003)	-	-

We next assessed the level of genetic differentiation between different hierarchical levels within the native and invasive ranges separately. We found significant genetic differentiation at most hierarchical levels in both ranges, although there were substantial differences in the magnitudes of differentiation in the native and invasive habitats. In the native range, measures of Θ were relatively low (Table 2-3). In contrast, all measures of Θ for the invasive range were high and strongly significant (Table 2-3). Overall, there was substantially more genetic differentiation across hierarchical levels in the invasive range than the native range.

Pairwise F_{st} values were calculated between all regions (Table 2-4). Values of F_{st} were often less than 0.01 for comparisons within the native region, indicating relatively low levels of differentiation. In contrast, values of F_{st} in the invasive range were substantially higher, with many estimates of Fst among the Hawaiian regions substantially greater than 0.1. In addition, all pairwise comparisons involving Kauai had F_{st} values

greater than 0.2, suggesting that Kauai may be the most genetically distinct island in the invasive range.

Table 2-4 Pairwise estimates of F_{st} for all regions.

	California	Colorado	Oregon	Wyoming	New Mexico	Molokai	Hawaii	Kauai	Lanai	Maui
Colorado	0.009*									
Oregon	0.005*	0.008*								
Wyoming	0.019*	0.005	0.014							
New Mexico	0.029*	0.026*	0.034*	0.031*						
Molokai	0.083*	0.095*	0.081*	0.111*	0.107*					
Hawaii	0.033*	0.046*	0.034*	0.054*	0.072*	0.124*				
Kauai	0.171*	0.197*	0.168*	0.244*	0.277*	0.244*	0.222*			
Lanai	0.091*	0.101*	0.095*	0.122*	0.101*	0.095*	0.118*	0.279*		
Maui	0.049*	0.059*	0.053*	0.076*	0.074*	0.119*	0.084*	0.230*	0.067*	
Oahu	0.064*	0.084*	0.066*	0.108*	0.109*	0.096*	0.108*	0.218*	0.135*	0.111*

* $p < 0.05$

We tested for the presence of genetic isolation by distance. Our analysis revealed no significant correlation between genetic distance (F_{st}) and geographic distance (km) for individuals sampled from different traps in the native range (Mantel test r = 0.042, p = 0.102) (Figure 2-2a). In contrast, there was a strong and significant isolation by distance relationship between individuals sampled from different traps within the invasive region (r = 0.569, p < 0.001) (Figure 2-2b). We found no evidence for isolation by distance within the individual Hawaiian Islands of Kauai (r = -0.011, p = 0.546), Oahu (r = 0.062, p = 0.181), Molokai (r = -0.028, p = 0.665), or Lanai (r = 0.120, p = 0.157). However, we did find significant genetic isolation by distance relationships in Hawaii (r = 0.163, p = 0.005) and Maui (r = 0.063, p = 0.002).

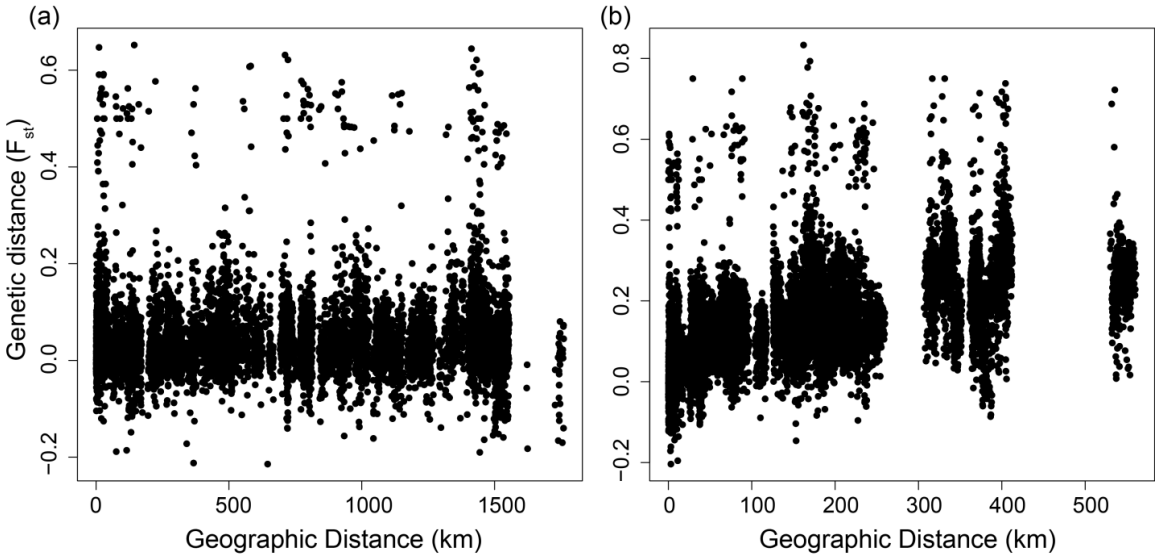

Figure 2-2 Relationship between genetic distance (F_{st}) and geographic distance (km) in the (a) native mainland (Mantel test r = 0.042, p = 0.102) and (b) invasive Hawaiian range (r = 0.569, p < 0.001) of *V. pensylvanica*.

2.4.3 Genetic clustering

V. pensylvanica wasps from the native and invasive ranges were both clustered into putative populations based on their multilocus genotypes. This analysis grouped individuals into two genetically distinct populations (Figure 2-3a). All of the individuals from the native range formed a single population. Individuals from the islands of Hawaii, Kauai, and Oahu in the invasive regions clustered into this single population. Conversely, individuals from Molokai, Lanai, and Maui formed another distinct population separate from the other island and the mainland regions. Most individuals were assigned to a single cluster, suggesting a lack of admixture.

We performed the clustering analysis considering only individuals from the native range. In this case, it was difficult to assign the most likely number of populations (K), as the ΔK metric was similar for K = 2 and 3. Nevertheless, K = 3 had the highest ΔK. Interestingly, almost all individuals in the native range were partially assigned to all three putative populations (Figure 2-3b). Individuals sampled from Balboa Park transect in California were a slight exception and tended to form a more distinct group than individuals sampled from other transects. However, the overall analysis indicated a general lack of genetic structure within the native range of *V. pensylvanica*.

We next clustered individuals within the invasive range only. The ΔK metric produced a clear peak, in contrast to the analysis of the native range samples, suggesting the most likely number of populations was seven (Figure 2-3c). Individuals within islands tended to form distinct clusters. Specifically, *V. pensylvanica* within Molokai, Kauai, and Lanai each formed separate and distinct populations. Samples from Oahu were separated into two clusters, where individuals were either part of a cluster that also consisted of samples from Hawaii or part of a cluster that consisted solely of samples from Oahu (Figure

2-3c). All of the Oahu individuals that clustered with Hawaiian samples were collected from a single transect, Satellite Road, and showed no signs of admixture. In contrast, individuals from Maui displayed signs of admixture. Most individuals from Maui were partially assigned to two clusters; however, the fractional memberships of individuals varied by transect. Individuals from the transects of Hosmer Grove, Haleakala, Olinda Road, and Waipoli Road had a higher probability of being assigned to one cluster (Figure 2-3c) while individuals from Maui Iao Valley and Waihee Ridge Trail had a higher probability of being assigned to the other Maui cluster (Figure 2-3c). Regardless, the overall analysis of samples from the invasive range showed substantial evidence for population genetic structure both within and between islands.

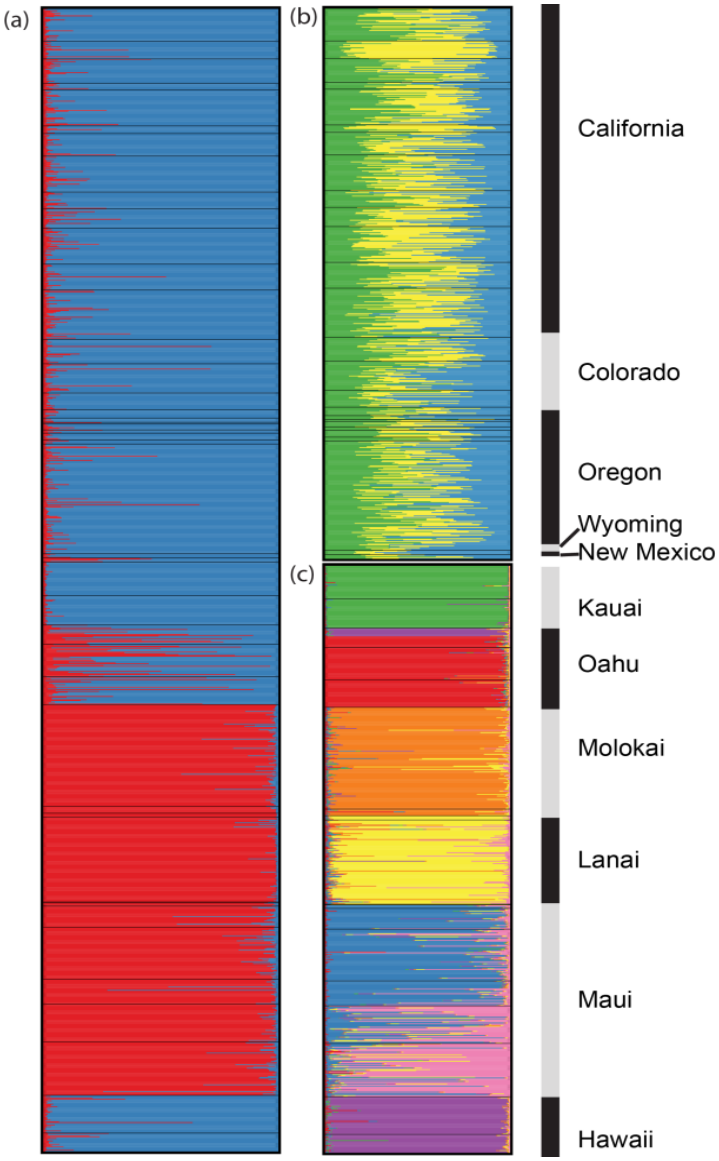

Figure 2-3 Estimated membership coefficients for individuals in each of K putative populations in the (a) combined native and invasive ranges (K = 2), (b) native range only (K = 3), and (c) invasive range only (K = 7). All transects in Hawaii are ordered west to east starting at the top with the island of Kauai. Each line represents an individual, the color of which corresponds to the estimated membership of that individual in a certain cluster; the same colors are used to represent different populations in the different figure panels. Sample origin is denoted by grey and black bars.

We assigned individuals from the invasive range to regions in the native range to determine the potential origins of invasive *V. pensylvanica* in Hawaii. We found that 80%

of the individuals from the invasive range had the highest score of being assigned to the western regions of the native range (Table A 3a). Overall, the mean assignment scores for the western regions were higher than those found for the central regions (Table 2-5a). This suggested that the western part of the native range was the most probable source of the invasive population. Notably, however, exclusion probabilities were generally high for all individuals ($P > 0.05$), suggesting that we cannot exclude either the western or the central regions as the source population for invasive *V. pensylvanica* (Table A 3b).

In contrast to the assignment tests, approximate Bayesian computation suggested that the central regions of the native range were the most likely source of the invasive populations (Table 2-5b). Specifically, the scenario where individuals from the invasive range were derived from the central part of the native range without a bottleneck had the highest probability among different tested scenarios (Posterior Probability = 0.655; 95% C.I. of 0.603 to 0.708).

Table 2-5 Assignment of invasive *V. pensylvanica* populations to the western regions (California and Oregon) or central regions (Wyoming, Colorado, and New Mexico) of the native range. (a) Assignment scores of individuals from invasive regions to combined reference regions. (b) Relative posterior probability (with 95% C.I. in parentheses) for demographic scenarios where invasive regions were derived from either the western or central regions with or without associated bottlenecks.

	a. Assignment Scores		b. Relative Posterior Probability			
Island	**West**	**Central**	**West, no bottleneck**	**Central, no bottleneck**	**West, bottleneck**	**Central, bottleneck**
Kauai	90.2	9.8	0.28 (0.149 – 0.410)	0.504 (0.410 – 0.598)	0.07 (0.000 – 0.173)	0.146 (0.066 – 0.226)
Molokai	67.5	32.5	0.06 (0.038 – 0.082)	0.735 (0.683 – 0.786)	0.01 (0.000 – 0.022)	0.195 (0.148 – 0.243)
Maui	76.2	23.8	0.142 (0.095 – 0.190)	0.706 (0.648 – 0.764)	0.017 (0.008 – 0.026)	0.134 (0.098 – 0.171)
Lanai	70.2	29.8	0.142 (0.100 – 0.185)	0.726 (0.673 – 0.779)	0.019 (0.010 – 0.027)	0.113 (0.082 – 0.143)
Hawaii	87.6	12.4	0.27 (0.210 – 0.329)	0.575 (0.515 – 0.635)	0.031 (0.012 – 0.050)	0.125 (0.091 – 0.158)
Oahu	90.9	09.1	0.122 (0.085 – 0.160)	0.78 (0.733 – 0.825)	0.018 (0.011 – 0.024)	0.081 (0.058 – 0.105)
All Islands	78.2	21.8	0.283 (0.194 – 0.371)	0.307 (0.217 – 0.398)	0.199 (0.120 – 0.277)	0.211 (0.131 – 0.291)

We visualized the relationships between individuals sampled from different transects with neighbor-joining trees. Transects from the native regions formed a star-like structure, indicating a lack of strong genetic relationships in the native range (Figure 2-4a). In contrast, transects in the invasive range from the same island clustered together, reflecting the genetic differences between islands (Figure 2-4b). We also produced a neighbor-joining tree of all regions in both the native and invasive ranges and found that samples from Maui, Lanai, and Molokai formed a single group while samples from Hawaii, Oahu, and Kauai grouped with mainland regions (Figure 2-4c).

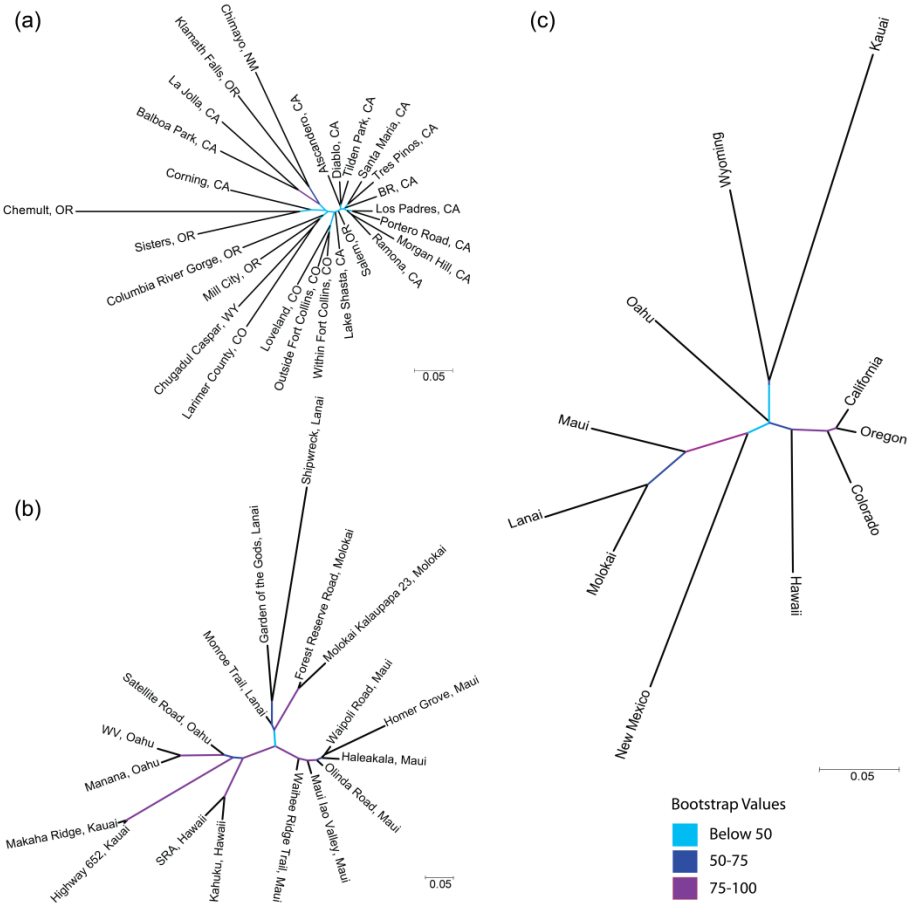

Figure 2-4 Unrooted neighbor joining trees for (a) native transects only (b) invasive transects only, and (c) all regions. Bootstrap support for nodes is represented by color.

2.5 Discussion

2.5.1 Small reduction in genetic diversity in the invasive range of V. pensylvanica

Introduced species tend to experience drops in genetic diversity due to population bottlenecks derived from founder events (Luikart et al. 1998; Goodisman et al. 2001; Sakai et al. 2001; Dlugosch and Parker 2008). Reduced genetic diversity could have negative effects on invasion success because it can affect a population's growth and ability to adapt to changing selection pressures (Sakai et al. 2001). However, a lack of genetic diversity does not necessarily preclude population growth or adaptation (Dlugosch and Parker 2008; Moran and Alexander 2014). Thus there is considerable interest in understanding if invasive species experience losses of genetic diversity and whether such losses are associated with invasion success (Dlugosch and Parker 2008; Purcell et al. 2012; Moran and Alexander 2014).

We compared levels of genetic diversity found within the native and invasive ranges of *V. pensylvanica*. Overall, greater levels of genetic diversity were observed in the native range than the invasive range (Table 2-2). However, the differences in variation were modest. *V. pensylvanica* from the invasive range had 97% of the expected heterozygosity and 64% of the allelic richness found in the native range. The overall drop in expected heterozygosity is quite small compared the drop in allelic richness, which is a characteristic of a brief population bottleneck (Luikart et al. 1998). In this case, some rare alleles are lost, although observed heterozygosity, which is more strongly influenced by common alleles, is not severely reduced (Luikart et al. 1998).

The allelic richness lost by invasive *V. pensylvanica* is similar to that lost by some other invasive social insects, such as the Formosan subterranean termite, *Coptotermes formosanus*, and the paper wasp, *Polistes chinensis antennalis*, in their invasive ranges (Husseneder et al. 2012; Tsuchida et al. 2014). In contrast, the Argentine ant, *Linepithema humile*, the Eastern Subterranean termite, *Reticulitermes flavipes*, and the Buff-tailed bumblebee, *Bombus terrestris*, experienced drops in allelic richness of 50% or more in their invasive ranges (Tsutsui et al. 2000; Vargo 2003; Schmid-Hempel et al. 2007).

In addition to a reduction of genetic diversity in the invasive range, we detected some evidence of genetic bottlenecks in the eastern islands of Molokai, Maui, Hawaii, and Lanai. Interestingly, there was no significant evidence for bottlenecks in Kauai and Oahu. The populations on both of these islands were introduced in the early 1900's (Nakahara 1980), so it is possible that allelic diversity and heterozygosity may have reached equilibrium, making it difficult to detect bottlenecks (Cornuet and Luikart 1996; Luikart et al. 1998). In contrast, populations on Molokai, Maui, Hawaii, and Lanai were established more recently and may not have reached equilibrium with respect to allelic heterozygosity.

In other invasive social insects, losses of genetic diversity have been implicated in the development of supercolonies, which are large, multi-queen colonies that consist of multiple nests and lack substantial boundaries (Holway et al. 2002; Tsutsui and Suarez 2003; Suarez and Tsutsui 2008; Helantera et al. 2009). Hanna et al (2014) showed that workers from native colonies of *V. pensylvanica* were always produced by a single queen, whereas colonies in the invasive range often contained workers produced by multiple queens (Goodisman et al. 2001; Hanna et al. 2014a). It is possible that the reduction of genetic diversity found in invasive *V. pensylvanica* is associated with this change in social

structure and invasion success. However, the magnitude of genetic diversity in the invasive range is still high compared to other introduced species that produce supercolonies (Helantera et al. 2009). In addition, perennial *V. pensylvanica* colonies can be found in parts of the native range, suggesting that phenotypic plasticity, rather than changes in genetic diversity, might be involved in the formation of multi-queen, perennial *Vespula* nests (Gambino 1991; Visscher and Vetter 2003).

2.5.2 Lack of genetic structure in the native range of V. pensylvanica

Vespula pensylvanica showed a remarkable lack of genetic structure in its native range in the United States, which stands contrary to our original prediction (Table 2-3). There was little evidence for genetic differentiation among hierarchically sampled locations and no significant evidence of genetic isolation by distance, suggesting a high level of gene flow across the entire native range (Figure 2-2a and Table 2-3). This is particularly notable because our sampling scheme spanned over 2000 km of western North America. The lack of genetic structure in the native range of *V. pensylvanica* could have resulted from human-mediated dispersal, which may have led to high rates of gene flow across the native range (Moller 1996). Alternatively, the dispersal distances of *Vespula* queens may be sufficient to develop genetic homogeneity over long, evolutionary timescales.

Our finding that native *V. pensylvanica* lacks genetic structure parallels results found in other native Vespula species over smaller ranges. For example, Hoffman et al (2008) failed to detect genetic structure in *Vespula maculifrons* and *Vespula squamosa*

along a span of approximately 130 km in its native range of North America (Hoffman et al. 2008). The sampling range for *V. maculifrons* and *V. squamosa* was more than an order of magnitude smaller than that of *V. pensylvanica* in this study, yet the results showing genetic homogeneity of Vespula species in their native ranges are consistent in these studies. However, the patterns seen in Vespula species are contrary to those found in other invasive Hymenoptera, which tend to display more population structure in their native range than their invasive range (Auger-Rozenberg et al. 2012; Tsuchida et al. 2014).

2.5.3 The invasion history of V. pensylvanica in Hawaii

We attempted to identify potential source populations of invasive *V. pensylvanica* in Hawaii. Given historical records, we expected that Oregon would be the source of invasive populations (Nakahara 1980). In accord with this prediction, we found that a majority of introduced individuals were assigned to the western part of the native range using one particular assignment test (Table 2-5A). However, a different assignment procedure suggested that the central area of the native range was the most likely source of invasive *V. pensylvanica* (Table 2-5B). These contrasting results may reflect the general lack of genetic differentiation among native *V. pensylvanica* populations, which may make the determination of the source of invasive populations difficult to ascertain. In addition, our limited sampling from the central regions of native *V. pensylvanica* may preclude our ability to assign source populations with high confidence (Muirhead et al. 2008).

V. pensylvanica displayed substantial levels of genetic differentiation both between and within Hawaiian Islands, in contrast to our expectations that genetic differentiation

would be limited. For example, we found that the relationship between genetic and geographic distance in the invasive range was nonlinear and displayed gaps at certain spans of geographic distance (Figure 2-2b). This trend mostly reflected strong genetic differences between islands combined with modest genetic structure within islands. The difference in patterns of isolation by distance between the native and invasive ranges was particularly notable given the great difference in geographic distance in the ranges. The native range stretches across 2000 km of western United States, while the invasive range spans approximately 600 km. Yet the invasive range showed substantially greater levels of genetic structure and isolation by distance. These trends could result, in part, from the expanses of ocean between the Hawaiian Islands (Pierce et al. 2014). The differences between native and introduced species may also reflect non-equilibrium conditions found in the introduced range (Akre et al. 1981). A similar, significant isolation by distance relationship, spanning approximately 225 km, was found for *V. germanica* in its invasive region of Australia (Goodisman et al. 2001).

We also found significant differences in the levels of genetic diversity among the Hawaiian Islands. Out of all the islands, Kauai had the lowest levels of all diversity metrics (Table 2-2). A survey conducted by the Hawaii Department of Agriculture suggests that *V. pensylvanica* was introduced multiple times to Kauai (Nakahara 1980). Multiple introductions are generally expected to result in greater genetic variation in an invasive habitat (Sakai et al. 2001; Kolbe et al. 2004). Therefore, it was somewhat unexpected that Kauai would have low levels of genetic variation. Kauai was also the most genetically distinct population when compared to other islands in the invasive range. It is possible that

a bottleneck may have led to this genetic differentiation. Alternatively, Kauai may have been colonized by *V. pensylvanica* populations from an un-sampled region.

The *V. pensylvanica* population on Oahu was also thought to have been founded by multiple introductions (Nakahara 1980). However, levels of genetic diversity in Oahu were relatively high and individuals within Oahu formed two distinct populations (Figure 3a). These results suggest that *V. pensylvanica* may have been introduced multiple times to Oahu. This island contains a large percent of the Hawaiian Island's human population so it is possible that such putative introductions were facilitated by human mediated shipments from the mainland (Nakahara 1980).

The existence of a discrete genetic population consisting of all individuals from Hawaii and a few individuals from Oahu was surprising since Oahu and Hawaii are separated by the islands of Molokai, Lanai, and Maui (Figure 2-1, Figure 2-3c). It thus seems unlikely that Oahu individuals directly seeded the Hawaiian population, or vice versa, through natural dispersal. However, human-mediated dispersal could account for this pattern. Regardless, all of the aberrant Oahu individuals originated from a single transect that was the most western of all the Oahu transects. Since these individuals were confined to the western section of the island, physical barriers, like the volcanoes Wai'anae or Ko'olau, may have prevented gene flow and the homogenization of allele frequencies across Oahu (Roderick and Gillespie 1998).

The islands of Molokai and Lanai also each formed a separate, genetically distinct population from all of the other islands (Figure 2-3c). This is consistent with the idea that Molokai and Lanai were colonized through single, separate introductions in the late 1970s

(Nakahara 1980). Even though both islands formed a distinct population, Molokai and Lanai cluster together in the NJ tree (Figure 2-4b). The populations on both islands may have been seeded by genetically similar source populations. Substantial genetic drift may have occurred during population formation, creating genetically distinct populations on each island.

Finally, we uncovered an unusual pattern of genetic structure in Maui. Individuals from transects on the eastern part of Maui, tended to form a somewhat differentiated population from those on the western part of the island (Figure 2-3c). This suggests low levels of gene flow between western and eastern *V. pensylvanica* in Maui, possibly to due to physical isolation (Roderick and Gillespie 1998). Such a result also raises the possibility that there may have been at least two introductions of *V. pensylvanica* to this island.

2.5.4 Conclusions

We examined the population genetic structure of *Vespula pensylvanica*, a wasp introduced to the archipelago of Hawaii from its native range in North America (Nakahara 1980; Gambino and Loope 1992). Remarkably, we found that invasive populations displayed substantially higher levels of genetic structure than native populations. Thus, *V. pensylvanica* is capable of high levels of dispersal and gene flow, likely through human-mediated transportation. However, such gene flow is apparently constrained in the invasive habitats of Hawaii, which consists of islands separated by wide expanses of water. The presence of genetic structure in invasive populations reflects the influence of geographic barriers, invasion dynamics, and a non-equilibrium state of population structure. Continued

study of this taxon over the coming decades may be particularly useful for understanding how invasive species come to be established in their introduced habitats. Overall, study of the invasion of *V. pensylvanica* in Hawaii may provide further insight on the process of biological invasions on archipelagos, which could help in the development of policy that can prevent and curb invasions to vulnerable regions.

Acknowledgements

We would like to thank Karl M. Glastad, Micaela O. Alcala, and Amy Groh for their comments on manuscript. This research was partially supported by the Elizabeth Smithgall Watts and the Bennie H. and Nelson D. Abell Endowment Fund.

CHAPTER 3. GENETIC DIVERSITY AND DIFFERENTIATION OF NAKED MOLE RATS, *HETEROCEPHALUS GLABER*, IN ZOO POPULATIONS

3.1 Abstract

The naked mole rat, *Heterocephalus glaber*, is a highly unusual mammalian species that displays a complex social system similar to that found in eusocial insects. Colonies of *H. glaber* are commonly maintained in zoo collections because they represent fascinating educational exhibits for the public. However, little is known about the genetic structure of captive populations of *H. glaber*. In this study, we developed a set of microsatellite markers to examine genetic variation in three captive zoo populations of *H. glaber*. We also studied sex ratio of these captive populations. Our goal was to determine levels of genetic variation within, and genetic differences between, captive populations of *H. glaber*. Overall, we found modest levels of genetic variation in zoo populations. We also found little evidence for inbreeding within the captive populations. However, there was some evidence of genetic differentiation across the zoo populations, which may reflect the isolation of captive naked mole rat colonies. Finally, we found no evidence of biased sex ratios within colonies. Overall, our study is one of the first to document levels of genetic variation and sex ratios in a captive eusocial mammalian population. Our results may provide insight into how to manage captive populations of *H. glaber*.

3.2 Introduction

Eusocial species live in highly developed and interdependent societies (Wilson and Hölldobler 2005). For example, eusocial insects, like ants and termites, are known for their division of labor in which individuals are tasked with specific jobs in the colony (Wilson 1990). In general, some members of a colony reproduce (i.e., queens and kings) while others (workers and soldiers) aid in the care of the reproductive members and their offspring. Such a social system has been of interest for evolutionary biologists because some individuals forgo personal reproduction in order to aid the reproduction of family members. This social system poses challenges for models of natural selection which consider direct selection only. However, Hamilton developed the theory of kin selection, which considered both direct and indirect effects of selection, to explain the evolution and benefits of eusocial behavior (Hamilton 1964; Queller and Strassmann 1998; West et al. 2007).

In addition, kin selection has implications for the evolution of sex ratios in eusocial species (Trivers and Hare 1976; Boomsma and Grafen 1991; Bourke 2015). Fisher's sex ratio theory suggests that an equal sex ratio arises when the fitness returns to the group controlling sex investment are equal when raising either a male or female (Fisher 1930). Sex allocation theory also predicts that parents will adjust the sex ratios of their offspring according to their ability to invest in a specific sex and the resulting profit of that sex to the parent (Trivers and Willard 1973). Therefore, sex ratios can also be tied to the condition of the parents in species where the sexes have different future reproductive success. Interestingly, the genetic structure and interactions of eusocial species may lead to unusual sex ratios. Most investigations of sex ratio in eusocial species have focused on the eusocial Hymenoptera, which include eusocial ants, bees, and wasps (Trivers and Hare 1976;

Boomsma and Grafen 1991). Therefore, little is known about sex ratios in other eusocial species.

The naked mole rat, *Heterocephalus glaber*, is one of the only eusocial vertebrates (Jarvis 1981). *H. glaber* is a unique mammal, known for its odd appearance and unusual social behaviors. *H. glaber* are hairless, long-lived rodents with minimal sight capabilities (Sherman et al. 1992). Naked mole rats are native to Kenya, Ethiopia, Somalia, and Djibouti (Jarvis 1981) and live within large subterranean colonies. These colonies are headed by a reproductive caste that is responsible for the production of new offspring. However, most colony functions, such as foraging and nest maintenance, are undertaken by a separate nonreproductive caste. Both males and females remain in their natal colony, though some males do disperse (Oriain et al. 1996). Despite being one of the only eusocial vertebrates, little is known about the genetic structure and sex ratios of naked mole rats, particularly in captive populations.

A few previous studies have examined the population biology and genetics of wild *H. glaber* in Africa. One of the first studies examining the genetic structure of *H.glaber* found high levels of genetic similarity and inbreeding within colonies, leading to the widely accepted hypothesis that inbreeding and low dispersal rates drove the evolution of eusociality in this species (Reeve et al. 1990; Faulkes et al. 1997a). However, more recent studies have uncovered evidence for outbreeding in wild populations (Oriain et al. 1996; Braude 2000; Ciszek 2000). In addition, levels of genetic diversity in *H. glaber* populations may be higher than originally proposed (Ingram et al. 2015).

The goal of this study was to understand the population structure and sex ratios of captive populations of *H. glaber* within zoos. We developed a set of polymorphic DNA microsatellite markers to genotype individuals from three zoos across the United States. We also used genetic and morphological analyses to determine the sex of individuals from these colonies. Overall, this project provides insight into the effect of captivity in shaping the genetic structure and sex ratios of eusocial animals. Our study also has implications for captive breeding programs of these unusual animals (Earnhardt et al. 2001; Ivy and Lacy 2012; Lacy 2013).

3.3 Methods

3.3.1 Sample genotyping

We assayed the multilocus genotype of a total of 89 *H. glaber* individuals sampled from three zoos including Zoo Atlanta (ZA, 2013 – 2014; $n = 60$), San Diego Zoo (SDZ, 2006 – 2013, $n = 11$), and the Smithsonian National Zoological Park (SNZP, 2015, $n = 18$). Individuals from Zoo Atlanta were a part of a single colony, transferred from Houston Zoo, which received their colony from Point Defiance Zoo in Tacoma, Washington. The San Diego Zoo samples were derived from four in-house colonies which were originally initiated from colonies in the Philadelphia Zoo. The population from Smithsonian National Zoological Park also originated from the Philadelphia Zoo, with all but one individual coming from a single colony. All individuals analyzed in this study died of natural causes within their colonies and were then frozen at -80 C for subsequent analyses. DNA was

extracted from skin biopsies of individuals using a modified Chelex protocol (Goodisman et al. 2001).

New microsatellite primers for *H. glaber* were developed by analyzing the *H. glaber* genome v1.1 (Appendix Table 1) (Kim et al. 2011). Microsatellite regions were located using the program MISA (MIcroSAtellite) (Thiel et al. 2003). Flanking primer regions were developed using Primer3 v 2.3.7 (Untergasser et al. 2012) with parameters set as follows: product size between 100 and 1000 bp with the optimal size of 200bp, primer size between 18 and 27 bp with an optimal length of 20 bp, annealing temperature between 57 and 63 °C with an optimal annealing temperature of 60°C, and G-C content between 20 and 80%. Further probe information can be found in the NCBI Probe Database (Accession Pr032825906-Pr032825937). We also genotyped individuals with *H. glaber* primers developed by Ingram *et al.* 2014 (Hglab01, Hglab03, Hglab07, Hglab08, Hglab09, Hglab10, Hglab13, Hglab14, Hglab17, Hglab18, Hglab19, Hglab22) (Ingram et al. 2014).

Many loci were PCR-amplified with standard fluorescently-labelled primers (Hoffman et al. 2008). However, some loci were amplified using the M13-nested-PCR method (Schuelke 2000) (Table B 1). Regardless of the amplification method used, the resulting PCR amplicons were run on an ABI 3100 Genetic Analyzer (Applied Biosystems). Alleles were scored manually using GeneMapper (SoftGenetics). We initially genotyped all individuals at 44 microsatellite loci (Table B 2). However, two loci were not readily scored, so we removed them from the rest of the analyses, leaving 42 loci.

The sexes of individuals were determined genetically using a multiplex PCR assay (Katsushima et al. 2010). This protocol jointly amplifies part of the Y-linked DBY gene,

which serves to identify the sex of the target individual gene, and the 16S rRNA gene, which acts as a PCR-amplification control. Therefore, two DNA bands indicate male samples, while a single band appears for female samples. The sexes of some individuals were also determined by dissection. We used a χ^2 goodness-of-fit test to investigate if the sex ratios of the zoo populations deviated from 50:50.

3.3.2 Genetic analysis

Genetic diversity measures, including number of alleles (N_a), effective number of alleles (N_e), observed heterozygosity (H_o), expected heterozygosity (H_e), and inbreeding coefficient (F_{is}) were calculated with the program GenAlEx v 6.5(Peakall and Smouse 2012). Kruskal-Wallis Tests were performed to see if there were significant differences in these genetic diversity measures across the three zoo populations. The probability test of GENEPOP v 4.6 was used to test for deviations of genotype frequencies from Hardy-Weinberg equilibrium and for the presence of linkage disequilibrium among loci (Rousset 2008). The Bonferroni correction was used to adjust for multiple testing. The level of genetic differentiation between zoo populations was measured with Weirs and Cockerham's F_{st} using FSTAT v 2.9.3 (Weir and Cockerham 1984; Goudet 1995).

Individuals were grouped into putative populations, or genetic clusters, using the program STRUCTURE v 2.2 (Pritchard et al. 2000). The number of genetic clusters (K) present across all the zoos was identified using different simulations, each under the assumption of a different K value (1 to 3). To account for individuals with mixed ancestry, we used an admixture model with uncorrelated allele frequencies. We also utilized the

LOCPRIOR model to use sampling location to inform clustering. For each K, we ran each simulation 10 times with 10,000 steps of burn-in and 100,000 MCMC iterations. The most likely value of K was identified based on log likelihood and the ΔK method developed by Evanno et al. (2005) as implemented in STRUCTURE HARVESTER (Pritchard et al. 2000; Evanno et al. 2005; Earl and Vonholdt 2012). CLUMPP v 1.1.2 was then used to align the 10 replicate runs for each set of K simulations (Jakobsson and Rosenberg 2007). The clustering results were then visualized with Distruct v 1.1 (Rosenberg 2004).

3.4 Results

3.4.1 Genetic diversity

We examined the levels of genetic diversity of the ZA, SNZP, and SDZ *H. glaber* populations at 42 microsatellite loci. A total of 24 of those loci were monomorphic across all of populations (Table 3-1). Our initial analysis did not detect significant deviations from Hardy-Weinberg equilibrium in the SNZP ($\chi^2_{df = 22} = 14.494$; $p = 0.883$; Kruskal-Wallis Test) or the SDZ populations ($\chi^2_{df = 30} = 20.515$; $p = 0.902$). However, there was evidence for deviations from Hardy-Weinberg equilibrium in the ZA population ($\chi^2_{df = 26} = 66.242$; $p < 0.001$). In contrast, we found no evidence for significant linkage disequilibrium between any pair of loci in any population.

We compared the levels of genetic diversity between the zoo populations (Table 1). We found no significant differences in number of alleles ($\chi^2_{df = 3} = 3.605$, $p = 0.307$; Kruskal-Wallis Test), number of effective alleles ($\chi^2_{df = 40} = 41.58$, $p = 0.402$), observed

heterozygosity ($\chi^2_{df = 36}$ = 40.621, p = 0.274), or expected heterozygosity ($\chi^2_{df = 38}$ = 36.62, p = 0.533) among populations. Each population had a few private alleles: 1 in ZA, 4 in SNZP, and 7 in SDZ. Levels of F_{is}, which estimates the degree of inbreeding in a population, are presented in Table 1. There was a significant, negative F_{is} for the ATL population. F_{is} levels were small and not significant for both the SNZP and SDZ populations.

Table 3-1 Genetic diversity measures of *H. glaber* from Zoo Atlanta, Smithsonian National Zoological Park, and San Diego Zoo for genetically variable loci.

Microsatellite	Zoo Atlanta					Smithsonian National Zoological Park					San Diego Zoo				
Locus	N_a	N_e	H_o	H_e	F_{is}	N_a	N_e	H_o	H_e	F_{is}	N_a	N_e	H_o	H_e	F_{is}
Hgla_6757.2	2	1.806	0.569	0.446	-0.275	1	1.000	0.000	0.000	-	2	1.385	0.333	0.278	-0.200
Hgla_7804	2	1.636	0.528	0.389	-0.359	2	1.895	0.529	0.472	-0.121	2	1.984	0.364	0.496	0.267
Hgla_7221.2	1	1.000	0.000	0.000	-	1	1.000	0.000	0.000	-	2	1.980	0.700	0.495	-0.414
Hgla_2663	1	1.000	0.000	0.000	-	2	1.882	0.625	0.469	-0.333	2	1.342	0.300	0.255	-0.176
Hgla_4233.1	2	1.718	0.262	0.418	0.374	2	2.000	0.111	0.500	0.778	2	1.471	0.000	0.320	1.000
Hgla_6197	2	1.160	0.149	0.138	-0.080	2	1.220	0.200	0.180	-0.111	1	1.000	0.000	0.000	-
Hglab17	1	1.000	0.000	0.000	-	2	2.000	0.385	0.500	0.231	2	2.000	0.500	0.500	0.000
Hgla_9415	2	1.518	0.436	0.341	-0.279	2	1.205	0.188	0.170	-0.103	2	1.976	0.667	0.494	-0.350
Hglab03	2	1.552	0.463	0.356	-0.301	3	1.947	0.625	0.486	-0.285	2	1.753	0.375	0.430	0.127
Hglab07	1	1.000	0.000	0.000	-	2	1.064	0.063	0.061	-0.032	3	2.800	0.429	0.643	0.333
Hglab08	1	1.000	0.000	0.000	-	2	1.080	0.077	0.074	-0.040	2	1.280	0.250	0.219	-0.143
Hglab14	2	1.494	0.418	0.331	-0.264	2	1.074	0.071	0.069	-0.037	2	1.960	0.571	0.490	-0.167
Hglab18	2	1.625	0.519	0.384	-0.351	1	1.000	0.000	0.000	-	2	1.508	0.429	0.337	-0.273
Hglab09	2	1.766	0.600	0.434	-0.383	3	1.471	0.313	0.320	0.024	3	2.418	0.667	0.586	-0.137

Table 3-1 continued

Microsatellite	Zoo Atlanta					Smithsonian National Zoological Park					San Diego Zoo				
Locus	**N_a**	**N_e**	**H_o**	**H_e**	**F_{is}**	**N_a**	**N_e**	**H_o**	**H_e**	**F_{is}**	**N_a**	**N_e**	**H_o**	**H_e**	**F_{is}**
Hglab10	2	1.791	0.415	0.442	0.061	2	1.936	0.455	0.483	0.060	1	1.000	0.000	0.000	-
Hglab13	3	2.062	0.550	0.515	-0.068	2	1.600	0.500	0.375	-0.333	3	1.815	0.286	0.449	0.364
Hglab19	2	1.999	0.380	0.500	0.240	1	1.000	0.000	0.000	-	3	2.323	0.500	0.569	0.122
Hglab22	2	1.975	0.472	0.494	0.044	2	1.912	0.357	0.477	0.251	4	3.459	0.500	0.711	0.297
MEAN	1.333	1.217	0.137	0.124	-0.126	1.381	1.197	0.107	0.110	-0.004	1.524	1.368	0.164	0.173	0.041
S.E.	0.081	0.055	0.034	0.030	0.038	0.090	0.055	0.030	0.029	0.044	0.119	0.091	0.037	0.037	0.055

N_a, observed number of alleles; N_e, effective number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{IS}, inbreeding coefficient index.

3.4.2 Genetic differentiation between colonies

We performed allelic probability tests for each population pair across all loci in Genepop to determine if the allele frequencies differed among the zoo populations. Each of the three pairwise comparisons between zoos was highly significant ($p < 0.001$). Pairwise F_{ST} was also measured between the zoo populations. F_{ST} estimates between ZA and SNZP was 0.498, between ZA and SDZ was 0.376, and between SDZ and SNZP was 0.446. Therefore, overall, we found high and significant measures of genetic differentiation between all zoo populations.

Analysis of the relationships among the zoo populations using the program STRUCTURE revealed the relationships among the three zoo populations (Figure 3-1). Our analyses suggested that the three zoo populations actually represented two distinct clusters (K = 2) (Figure 3-1A). Cluster 1 was composed of all individuals from the ZA population and Cluster 2 consisted of all individuals from the SNZP population. Surprisingly, there was the presence of admixture between the two clusters in the SDZ population. Interestingly, however, when we set the number of populations to K = 3, we recovered three clusters that corresponded to the three zoo populations (Figure 3-1B).

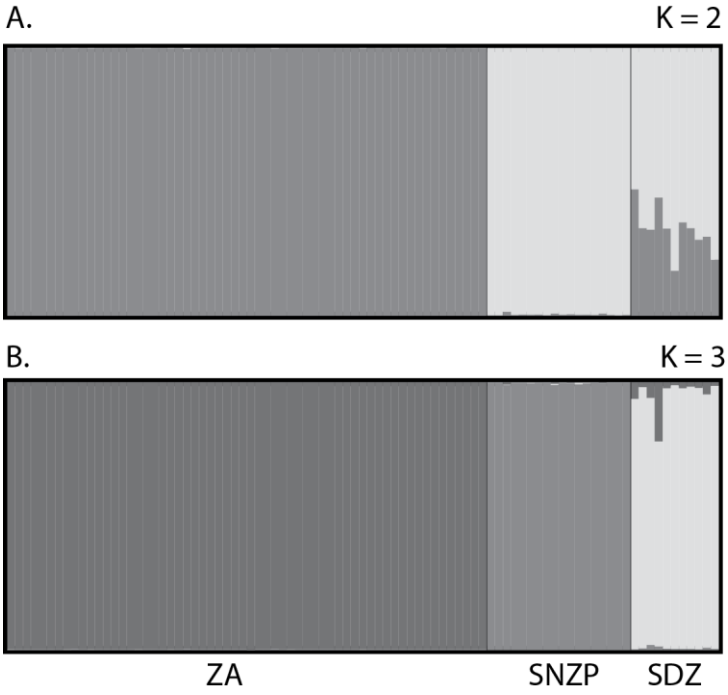

Figure 3-1 Estimated membership coefficients for naked mole rats from Zoo Atlanta (ZA), Smithsonian National Zoological Park (SNZP), and San Diego Zoo (SDZ) colonies identified using STRUCTURE. K represents the number of putative populations individuals are clustered into. Each line represents an individual, the shade of which corresponds to the estimated membership of that individual in a certain cluster. (A) K = 2. (B) K = 3.

3.4.3 Genetic sex identification

We determined the sex of many of the sampled individuals (Katsushima et al. 2010) (Table B 2). The ZA population consisted of 26 females and 34 males, and did not differ significantly from equality ($\chi^2_{df=1} = 1.067$; $p = 0.302$; χ^2 goodness-of-fit test). The sex ratios (f:m) of the SNZP and SDZ samples were 5:5 ($\chi^2_{df=1} = 0$; $p = 1$) and 6:4, respectively ($\chi^2_{df=1} = 0.4$; $p = 0.527$). Thus there was no evidence that sex ratio differed from equality in any of the captive populations.

3.5 Discussion

The goal of this study was to examine the levels of genetic variation and sex ratios of three captive colonies of naked mole rats. We sampled individuals from colonies at Zoo Atlanta (Atlanta, GA, USA; ZA), San Diego Zoo (San Diego, CA, USA; SDZ), and Smithsonian National Zoological Park (Washington, DC, USA; SNZP). We first analyzed the distribution of genetic variation within captive colonies to determine if genotype frequencies differed from Hardy-Weinberg equilibrium. We detected deviations from Hardy-Weinberg equilibrium in one (ZA) of our three populations. In particular, F_{is} for SNZP and SDZ were both not significant, whereas F_{is} for ZA was significant and negative. Therefore, there was an excess of heterozygotes in the ZA zoo population relative to expectations of random union of gametes. Such deviations are not unexpected and likely arise because colonies represent more or less complex families of related individuals. Therefore, sampled individuals are not genetically independent, as one would expect in a population in Hardy-Weinberg equilibrium.

Early genetic studies of natural populations of *H. glaber* suggested a high level of inbreeding for NMRs, which was believed to have facilitated the evolution of sociality in this species (Reeve et al. 1990). However, more recent studies suggest that the previously observed high level of inbreeding was an artifact of sampling bias. In particular, Ingram et al. (2015) investigated genetic structure of previously unstudied NMR populations, along with those that had been previously studied (Reeve et al. 1990; Faulkes et al. 1997a). Ingram et al. confirmed that the previously studied NMR populations had high inbreeding coefficients. However, the newly-analyzed populations had lower inbreeding coefficients (Ingram et al. 2015). Therefore, *H. glaber* colonies are apparently less inbred than

previously thought. When we compared the inbreeding coefficients of our zoo samples and those sampled from wild individuals in Ingram et al., we saw that the magnitudes of the inbreeding coefficient (F_{is}) from the zoo populations are on par with those from wild populations that were not inbred. Therefore, our study supports the hypothesis that naked mole rat colonies are not as inbred as previously suspected.

We detected population structure among the NMR zoo populations indicating that the zoo populations differ genetically. When we grouped individuals based off their multilocus genotype, we found that all zoo samples could be clustered into two groups. Individuals from ZA made up one cluster, individuals from SNZP made up the second cluster, and individuals from SDZ were admixed (Figure 3-1). Therefore, individuals from ZA and SNZP were most divergent. However, when we assumed that there were three clusters, all three populations could be differentiated.

Strong population structure, associated with geographic variation, was also found in *H. glaber* populations in its native range across Kenya (Ingram et al. 2015). The genetic differences between the zoo populations probably arose, in part, because the different zoo populations originated from different sources. ZA's naked mole rat colony originates from Point Defiance Zoo (Tacoma, Washington, USA) by way of Houston Zoo (Houston, Texas, USA). SNZP received their colony from Philadelphia Zoo in 1991. SDZ also received some samples from Philadelphia Zoo in 1992. This may explain why SDZ individuals seem to have shared ancestry with individuals from SNZP. It is thus notable that we were able to recover the relationships among zoo populations using genetic techniques. Also, the zoo colonies are spatially fragmented which restricts migration and gene flow leading to genetic differentiation of populations through drift over time (Frankham 2008).

We examined the sex ratio of each zoo population. We found no evidence that the sex ratio differed significantly from 50:50 in any population. Our sample sizes from the SDZ and the SNZP were quite small and therefore provided little power to detect significant deviations. However, the sample size from ZA was reasonable. So, overall, there is no evidence for sex-ratio bias within captive naked mole rat colonies. This result is in accordance with data from other mole rat species, which may suggest that the cost to produce each sex is equal (Fisher 1930; Begall and Burda 1998; Bennett and Faulkes 2000). However, there has been evidence for a male-biased sex ratio in the offspring of wild-captured giant mole-rats (*Fukomys mechowii*); though, this was not found amongst adults (Kawalika and Burda 2007). Interestingly, captive colonies of the giant mole-rat have a high female-biased neonate sex-ratio (Kawalika and Burda 2007). This difference between wild-caught and captive populations of *F. mechowii* could arise for multiple reasons; (1) the difference could be an artefact of captive breeding, (2) older males may have higher dispersal rates, leaving a lower number of adult males, or (3) older males may have less activity, leaving them less likely to be caught in the wild (Kawalika and Burda 2007). It will be interesting and important to determine if captive and wild populations of *H. glaber* also show differences in sex ratio.

Most research conducted on sex ratios in eusocial species has focused on haplodiploid eusocial insects (Trivers and Hare 1976; Boomsma and Grafen 1991; Queller and Strassmann 1998). In haplodiploid eusocial hymenopterans, the relatedness between sister workers is greater than between workers and their brothers. Since there is asymmetry in the level of relatedness between the two sexes, workers are predicted to invest more resources into the raising of sisters than brothers, which may lead to female-biased sex

ratios (Trivers and Hare 1976). Naked mole rats are diploid so there is no relatedness asymmetry. Inclusive fitness predicts a lack of sex ratio bias in these eusocial diploids. However, there are other mechanisms that could lead to adaptive sex ratio bias, such as local resource competition (Silk 1983), local resource enhancement, or helper repayment (Gowaty and Lennartz 1985; Emlen et al. 1986). Data in captive Damaraland mole-rat (*Fukomys damarensis*), another putatively eusocial rodent, has found evidence that partially supports each of these three mechanisms (Lutermann et al. 2014). Thus, factors such as sex-biased dispersal, differences in helping between the sexes, and the condition of the reproductives, possibly play a role in shaping the sex ratio of a colony. More theoretical work and extensive sampling must be done to further understand the association between eusociality and sex ratios in naked mole rats in both captive and wild populations.

Conclusions

1. Naked mole rats are fascinating vertebrates, which display a complex social system similar to that found in social insects.
2. We developed a set of microsatellite markers to examine levels of genetic variation in zoo populations of captive naked mole rats.
3. We found that captive naked mole rat populations displayed reasonable levels of genetic variation, substantial genetic differences, and relatively little inbreeding.
4. We found no evidence for sex ratio bias within the three zoo populations.
5. This research provides insight on the levels of genetic variation and sex ratio of captive naked mole rats, which may aid in the management and care of these interesting mammals.

Acknowledgments

We would like to thank Yemi Olubowale for laboratory assistance. This research was supported by the Elizabeth Smithgall Watts Fund.

CHAPTER 4. GENE DUPLICATION AND TH EVOLUTION OF PHENOTYPIC DIVERSITY IN INSECT SOCIETIES[2]

4.1 Abstract

Gene duplication is an important evolutionary process thought to facilitate the evolution of phenotypic diversity. We investigated if gene duplication was associated with the evolution of phenotypic differences in a highly social insect, the honeybee *Apis mellifera*. We hypothesized that the genetic redundancy provided by gene duplication could promote the evolution of social and sexual phenotypes associated with advanced societies. We found a positive correlation between sociality and rate of gene duplications across the *Apoidea*, indicating that gene duplication may be associated with sociality. We found that genes showing biased expression between *A. mellifera* alternative phenotypes tended to be found more frequently than expected among duplicated genes than singletons. Moreover, duplicated genes had higher levels of caste-, sex-, behavior-, and tissue-biased expression compared to singletons, as expected if gene duplication had facilitated phenotypic differentiation. We also discovered that duplicated genes were maintained in the *A. mellifera* genome through the processes of conservation, neofunctionalization, and specialization, but not subfunctionalization. Overall, we conclude that gene duplication may have facilitated the evolution of social and sexual phenotypes, as well as tissue

[2] Chau, L. M. and M. A. D. Goodisman. Gene duplication and the evolution of phenotypic diversity in insect societies.Evolution (in press).

differentiation. Thus this study further supports the idea that gene duplication allows species to evolve an increased range of phenotypic diversity.

4.2 Introduction

Individuals within species often belong to distinct phenotypic classes that have different functional roles. These classes (e.g., sexes) may experience contrasting selection pressures on traits associated with their distinct roles. Therefore, alleles favored in one class may be disfavored in the other if different classes share a majority of their genome. (Bonduriansky and Chenoweth 2009; Stewart et al. 2010; Pennell and Morrow 2013; Gotzek et al. 2015). Contrasting selection pressures may ultimately displace individuals of both classes from their phenotypic optima. Overall, this "intralocus conflict" represents a fundamentally important process inhibiting adaptation within species (Lande 1980; Rice and Chippindale 2001; Bonduriansky and Chenoweth 2009; Pennell and Morrow 2013; Rice 2013). Problems arising from intralocus conflict can be reduced through mechanisms that decouple the trait genetic correlation between the different phenotypic classes (Lande 1980). This allows each class to express different trait values in reaction to their contrasting selection pressures.

Gene duplication has been hypothesized to be a mechanism capable of relieving intralocus conflict (Ellegren and Parsch 2007; Connallon and Clark 2011; Gallach and Betran 2011). After a gene is duplicated, a pair of paralogs are created, each highly similar in sequence and redundant in function (Gu 2003). Such redundancy is thought to release a single paralog from selection after the duplication event, since there is an exact copy

retaining its original function (Ohno 1970; Lynch and Conery 2000). Mutation can then alter the function of the focal paralog, which will ultimately determine whether the gene pair is preserved in the genome either by genetic drift or positive selection (Proulx 2012; Cardoso-Moreira et al. 2016).

Gene duplicates are generally thought to undergo one of five possible outcomes within the genome: pseudofunctionalization, conservation, neofunctionalization, subfunctionalization, or specialization (Ohno 1970; Force et al. 1999; Lynch and Conery 2000; He and Zhang 2005; Innan and Kondrashov 2010). Most gene duplicates are expected to undergo pseudofunctionalization, which occurs when one paralog is silenced by mutations and becomes nonfunctional (Lynch and Conery 2000). However, there are circumstances that allow both paralogs to be functional and remain in the genome. Under conservation, the ancestral function of both paralogs is conserved because there is selective advantage for increased dosage (Ohno 1970). A paralog may also gain novel functions through the process of neofunctionalization (Ohno 1970). Alternatively, the function of the ancestral single copy gene may be divided amongst the two paralogs through subfunctionalization (Force et al. 1999). In this case, both paralogs are subjected to a loss of certain ancestral subfunctions. Thus, in order to maintain the function of the original, ancestral single-copy ortholog, both paralogs must be preserved (Force et al. 1999). Finally, specialization occurs when neofunctionalization and subfunctionalization work together, creating two copies that are distinct from each and the ancestral gene (He and Zhang 2005).

Social insects are interesting taxa in which to study the importance of gene duplication in the amelioration of intralocus conflict. These insects, which include the

social bees, ants, social wasps, and termites, are among the most dominant organisms on earth (Wilson 1990). The success of social insects arises, in part, from the caste system in which multiple distinct classes of individuals are responsible for completing different tasks within the colony (Wilson 1990).

Hymenopteran insect societies usually consist of three castes: queens, workers, and males. Queens and males are responsible for reproduction and dispersal. Workers perform tasks related to colony growth and maintenance, like brood care and foraging for food. Workers may be further subdivided into behavioral subcastes, such as nurses and foragers (Seeley 1982; Whitfield et al. 2003).The difference in behavior among the castes is often paired with drastic differences in morphology and physiology (Toth et al. 2010; Feldmeyer et al. 2014). Therefore, different castes experience strongly divergent selection pressures.

Importantly, hymenopteran social insect castes share a common genome (though males are haploid and female workers and queens are diploid) (Normark 2003). Thus genetic correlations between the castes can limit the evolution of caste dimorphism in reaction to divergence selection pressures (Gadagkar 1997a; Linksvayer and Wade 2005; Kovacs et al. 2010; Hall et al. 2013). Consequently, social insect castes may suffer a variety of intralocus conflicts, which may impede the elaboration of caste differences and limit the evolution of sociality.

The purpose of this study was to investigate if the genetic material provided by gene duplication ameliorated intralocus conflict, facilitating the specialization of social phenotypes within insect societies (Holman 2014). We hypothesized that the evolution of caste specialization was initially constrained by intralocus conflict. We further conjectured

that this conflict may have been lessened through the process of gene duplication (Gadagkar 1997a). Specifically, duplicated genes may have been co-opted in the development of different castes and thereby allowed the evolution of caste-specific function.

We investigated if gene duplication might be associated with the diversification of castes in the honey bee, *Apis mellifera*. *A. mellifera* societies contain standard queen, worker, and male castes, as well as nurse and forager worker behavioral subcastes. The presence of these alternative phenotypes, and the wealth of data on gene expression differences among castes (Whitfield et al. 2003; Zayed et al. 2012; Cameron et al. 2013; Elsik et al. 2014; Jasper et al. 2015; Ashby et al. 2016), makes *A. mellifera* an ideal system to study the role of duplication in the evolution of alternative phenotypes.

We studied the effects of gene duplication on castes in the honeybee using several approaches. First, we examined the relationship between the level of sociality and gene duplication across the *Apoidea* in order to determine if gene duplication was generally associated with the evolution of complex social behavior. Second, we explored differences in biased gene expression between duplicated genes (paralogs) and non-duplicated genes (singletons) within *A. mellifera*. We hypothesized that gene duplication would accelerate the rate of expression divergence between phenotypes by providing new copies of genes that could be co-opted in the development of differential expression. Therefore, we predicted that duplicates would be more likely to be differentially expressed between castes and sexes than singletons (Huminiecki and Wolfe 2004). Third, we examined expression divergence between duplicate pairs. We hypothesized that duplicates gained divergent functions amongst phenotypes (Connallon and Knowles 2005; Innocenti and Morrow

2010). Therefore, we expected to find divergent expression patterns between duplicated genes. Finally, we examined the evolutionary processes that maintained paralogs in the genome. We predicted that there would be a high proportion of duplicates that were maintained by processes that led to functional diversification like specialization, subfunctionalization, and neofunctionalization. Overall, this study provides new information on the role of gene duplication in the evolution of dimorphism, intralocus conflict, and sociality.

4.3 Methods

4.3.1 Identification of Duplicate Genes and Duplication Rates

We downloaded gene families from OrthoDB v9.1(Zdobnov et al. 2017), which has identified orthologs in a hierarchical fashion. We used custom perl scripts to parse gene duplicates that were duplicated in *A. mellifera* but were single-copy across *Apoidea*. We also identified novel duplicates in nine other species in Apoidea, *Apis florea*, *Bombus impatiens*, *Bombus terrestris*, *Eufriesea Mexicana*, *Durfourea novaengliae*, *Habropoda laboriosa*, *Lassioglossum albipes*, *Megachile rotundata*, and *Melipona quadrifasciata*, which vary in level of sociality (Kapheim et al. 2015). We determined species-specific duplication rates by incorporating divergence times from Cardinal et al 2013(Cardinal and Danforth 2013).

We examined the relationship between species-specific duplication rate and sociality independent of phylogenetic relationship using phylogenetic independent

contrasts (PICs) (Felsenstein 1985). PICs for species-specific duplicates per million years and sociality values were generated with the R package Analysis of Phylogenetics and Evolution (APE) (Paradis et al. 2004). This analysis relied on a phylogenetic tree and distances based off of Cardinal and Danforth (2013) and Kapheim *et al*. (2015): ((Hlab:91,((Mqua:68,(Bimp:13,Bter:13):55):10,(Emex:62,(Amel:19,Aflo:19):43):16):13) :15,Mrot:106):9, (Dnov:85, Lalb:85):30). The species were assigned sociality values based off of (Kapheim et al. 2015): 0 represented solitary species, 1 represented facultative basic eusocial, 2 represented obligate basic eusocial, and 3 represented complex eusocial species. The relationship between the level of sociality and rates of gene duplication per million years were then determined using the Spearman's rank correlation.

4.3.2 Gene Expression Data and Analysis

We investigated patterns of gene expression within *A. mellifera* to understand the relationship between gene expression and gene duplication. We obtained *A. mellifera* RNAseq reads from four different studies that investigated expression differences between *A. mellifera* female castes (queens and workers), sexes (workers and drones), worker behavioral states (nurses and foragers), and worker tissues (Cameron et al. 2013; Jasper et al. 2015; Ashby et al. 2016; Vleurinck et al. 2016). Ashby *et al.* analyzed expression differences between *A. mellifera* queen, worker, and drone whole body larvae at stage L5 (PRJNA260604) (Ashby et al. 2016). Similarly, Vleurinck *et al.* assessed caste and sex differences by investigating gene expression in the brains of *A. mellifera* queen, worker, and drone pupae (stages 4-5) (PRJNA193691) (Vleurinck et al. 2016). In contrast,

Cameron *et al.* studied expression in 60 hour (L3 larval stage) whole body *A. mellifera* queens and workers (PRJNA227348) (Cameron et al. 2013). Finally, Jasper *et al.* examined gene expression in *A. mellifera* adult worker nurses and foragers across 10 tissues: brain, antennae, midgut, hypopharyngeal gland, malpighian tubule, mandibular gland, muscle, nasonov gland, sting gland, and second thoracic ganglia (PRJNA243651 & PRJNA211831) (Jasper et al. 2015).

All RNAseq data were downloaded from NCBI's sequence read archive. The qualities of the raw RNA-Seq reads were assessed with FastQC v0.11.5 (Andrews 2010). Reads were then trimmed with Trimmomatic v 0.35 (Bolger et al. 2014). Adapters sequences were trimmed and low-quality bases were trimmed from either side of each read. A sliding window with a minimum quality score of 15 was applied to each read. RSEM 1.3.0 (RNA-Seq by Expectation Maximization) was used to measure expression levels (Li and Dewey 2011). RSEM was used with the Bowtie2 (version 2.2.2) aligner to align reads to the *A. mellifera* reference gene set (Amel OGSv3.2; http://hymenopteragenome.org/beebase/) (Langmead and Salzberg 2012). Expected read count was measured with RSEM with default settings. Bowtie2 within RSEM does not allow for indel, local, and discordant alignments, which may lead to lower alignment rates compared to Bowtie2 itself (Li and Dewey 2011). Also, the use of RSEM allows for the mapping of non-uniquely mapped reads that may have an impact on measuring the expression of duplicate genes. Details of the alignment procedure for each dataset are provided in Supplementary Table 1 (Table C 1).

Each RSEM file was concatenated into single dataset and differential expression of genes was determined with edgeR v 3.16.0 (Robinson et al. 2010). The trimmed mean of

M values (TMM) method was used for normalization of gene expression. Pairwise comparisons were made between castes (queens and workers), sexes (drones and workers), and behavioral states (nurses and foragers, brains only) to identify differentially expressed genes. The false discovery rate (FDR) was calculated using the Benjamini-Hochberg correction and a FDR less than or equal to 0.05 was considered significant (Benjamini and Hochberg 1995). Levels of differential expression were calculated as the absolute value of the $\log_2$ fold change between each pair. We calculated tissue expression specificity, τ, per gene across ten tissues (Yanai et al. 2005; Atallah et al. 2013; Jasper et al. 2015). Tau ranges from 0 to 1 with low values indicating that a gene is broadly expressed among tissues and high values indicating that a gene is expressed in few tissues.

We investigated if the frequency of different phenotype-biased genes differed between duplicated genes and singletons. Genes were classed into phenotype-biased gene categories (i.e., phenotypically-biased or phenotypically-unbiased) based off the FDR cut-off of 0.05 and expression direction (fold change equal or greater than two). Next, we used a chi-squared test to determine if the proportion of phenotype-biased genes depended on whether the genes were duplicates or singletons. Tests were conducted for caste-biased, sex-biased, and behavior-biased genes. Wilcoxon rank-sum tests were used to compare the levels of expression bias between duplicated genes and singletons.

Each pair of duplicate genes was then categorized based on the pair's joint patterns of expression bias. For example, both copies of a duplicated gene in a queen-worker comparison could show concordant expression, with both genes having the same expression bias (e.g., both queen-biased). Alternatively, the paralogs could show discordant expression patterns with one paralog being more highly expressed in one caste

than the other, or one paralog could be caste-biased and the other unbiased. The expected proportions of each paired class were generated by randomly sampling genes 10,000 times from the pool of duplicate pairs to create null distributions of paired genes (Mikhaylova et al. 2008; Wyman et al. 2012). The mean proportions generated from the null distribution provided the expected proportions of each class (Mikhaylova et al. 2008; Wyman et al. 2012). Chi-squared tests were then used to compare the observed proportions of gene pairs falling into each class to the expected proportions constructed from randomization approach.

Expression divergence between duplicate gene pairs was calculated for caste-, sex-, behavior-biased expression, and tau. This was calculated as the absolute value of $(x-y)/(x+y)$, with x being the expression measure in one paralog and y being the expression measure in the other. We then used Wilcoxon rank-sum tests to compare the level of expression divergence between duplicates on the same chromosome and different chromosome to determine if the location of duplicate genes in the genome was associated with gene expression divergence between duplicates.

We investigated if there was a correlation in expression bias for paralogs within duplicate pairs. This analysis determined if a paralog that showed highly biased expression in one phenotypic comparison (e.g., queens vs workers) also tended to show highly biased expression in another (e.g., males vs females). We then used the program JMP 11 to perform a multivariate analysis of variance (MANOVA) with the duplicate pair as the independent variable and all measures of phenotype-biased expression (i.e., caste-bias, sex-bias, etc.) as dependent variables. This analysis produced a partial correlation matrix that provided information on whether paralogs tended to show correlations in expression-bias.

4.3.3 Identifying models of duplicate gene maintenance

We used the methodology of Assis and Bachtrog (2013) to determine the processes that maintained duplicates in *A. mellifera*. Briefly, this method considers the relationships among multiple Euclidean distances between the expression profiles of a single copy ortholog in a closely related species, the expression profiles of both duplicate genes in the focal species, and the combined expression profile of the duplicates. Comparison of these expression distances provides insight into whether conservation, neofunctionalization, subfunctionalization, or specialization maintains the focal duplicate pair in the genome.

We determined the evolutionary processes maintaining duplicates in the *A. mellifera* genome (Assis and Bachtrog 2013) by analyzing genes that were duplicated in *A. mellifera* but single copy in the social bee, *Bombus terrestris*. Custom perl scripts were used to identify genes in the OrthoDB v9.1 database that were duplicated in *A. mellifera* with orthologs that were single copy in the rest of Apoidea lineage (Zdobnov et al. 2017). We used sequence similarity measures from BLAST to classify each *A. mellifera* paralog in a pair as the "D1" or "D2" copy (Assis and Bachtrog 2013; Wang et al. 2016). We used BLASTp to compare each paralog to the single copy ortholog in *B. terrestris*, using the e-value, identity, and alignment length as a measure of sequence similarity. D1 paralogs were those with higher sequence similarity (lower e-value, high identity, and long alignment length) to the *B. terrestris* ortholog whereas the D2 paralogs were those with lower sequence similarity (higher e-value, low identity, and shorter alignment length) to the ortholog. We generated the gene expression profiles for *B. terrestris* queens, workers, and

males at adult, larval, and pupal stages using the same methods previously provided for determining expression differences in *A. mellifera* (Harrison et al. 2015). We then determined the processes maintaining duplicates with the R package CDROM (Perry and Assis 2016).

4.3.4 Sequence Evolution of Duplicate Genes

We investigated patterns of sequence divergence of duplicate genes in order to examine how rates of sequence evolution differed between duplicate pairs. *A. mellifera* (OGSv3.2) duplicates and *B. terrestris* single copy orthologs (NCBI build 1.1) sequences were aligned using MACSE v1.02 (Ranwez et al. 2011). Gene trees were created under the assumption the duplicates were most closely related and the single copy ortholog was used as the outgroup. The codeml package within PAML (v4.7) was used to measure synonymous and nonsynonymous branch specific substitution rates of the duplicate genes (Yang 2007). All genes with dS > 3 were considered to be saturated with mutations and removed from the analysis.

4.4 Results

4.4.1 Duplication rates across the Apoidea

We identified the number of species-specific duplicates across different bee species within *Apoidea* (Figure 4-1). We then determined the rates of species-specific duplication events for each lineage. We found that *A. mellifera* had the highest rate of duplication at

6.1 duplicates per million years. In contrast, bees considered ancestrally solitary, such as *Dufourea novaeangliae* and *Megachile rotundata*, had rates lower than 0.4 duplications per million years. Overall, we observed a significant, positive correlation between the level of sociality and rate of species-specific duplication across the *Apoidea* (rho = 0.6566, p = 0.0392; uncorrected Spearman's correlation) suggesting that gene duplication might be associated with the evolution of sociality in bees. However, when we performed the analysis with the phylogenetic corrected level of sociality and rate of species-specific duplication, the correlation was no longer significant (rho = 0.5021, p = 0.1684, phylogenetically corrected Spearman's correlation).

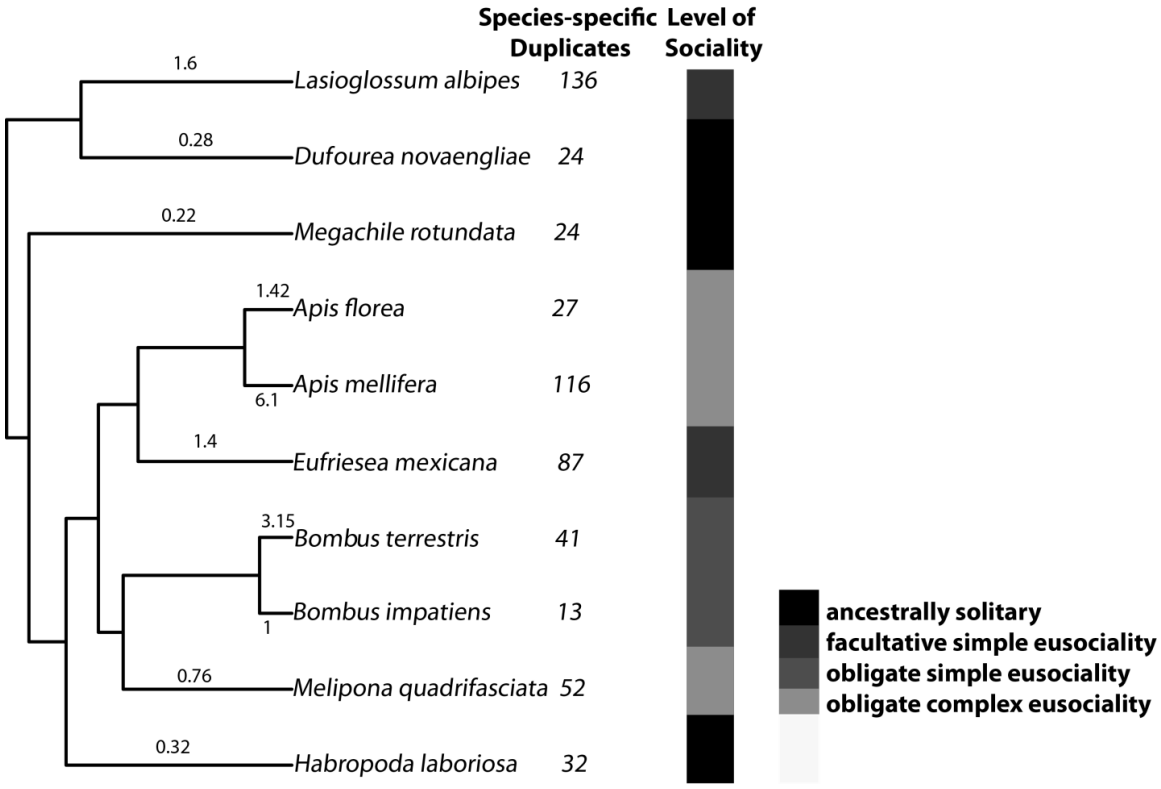

Figure 4-1 Species-specific duplications and duplication rates for different bee species varying in level of sociality. Numbers on branches represent species-specific duplication rates in duplicates/MY.

4.4.2 Differential expression between duplicates and singletons in A. mellifera

We identified 116 pairs of duplicated genes and 5235 singletons in *A. mellifera*. In order to further our understanding of the role of gene duplication in the evolution of alternative phenotypes, we examined the relationship between gene duplication and differential gene expression. First, we compared the proportions of caste-biased genes between duplicates and singletons (Table 4-1). Since there were a small number of duplicated genes showing biased expression, we performed chi-squared tests by grouping queen- and worker-biased genes into the overall category of "biased" genes. We found that there were significant differences in the percentage of caste-biased (i.e., queen- and worker-biased) genes between duplicated genes and singletons for two out of three datasets analyzed (Ashby *et al.*: $\chi^2_{df=1} = 2.14$, $p = 0.1435$; Vleurinck *et al.*: $\chi^2_{df=1} = 12.36$, $p = 0.0004$; Cameron *et al.*: $\chi^2_{df=1} = 5.24$, $p = 0.0220$, χ^2 test of independence). The patterns among datasets showed some similarities in that duplicated genes tended to show biased genes more often than expected (Table 4-1).

Table 4-1 Observed (and expected) counts of genes differentially expressed across castes for duplicated genes and singletons in three RNAseq datasets comparing queen and worker gene expression differences in A. mellifera.

Dataset	Expression	Duplicates	Singletons	Total
Ashby *et al*.	Queen-biased	15 (9.89)	243 (248.11)	258
NS	Unbiased	186 (191.43)	4808 (4802.57)	4994
	Worker-biased	3 (2.68)	67 (67.32)	70
	Total	204	5118	5322
Vleurinck *et al.*	Queen-biased	5 (0.68)	13 (17.32)	18
***	Unbiased	177 (188.74)	4853 (4841.26)	5030
	Worker-biased	18 (10.58)	264 (271.42)	282
	Total	200	5130	5330
Cameron *et al*.	Queen-biased	3 (1.17)	29 (30.83)	32
*	Unbiased	179 (183.22)	4830 (4825.78)	5009
	Worker-biased	4 (1.61)	40 (42.39)	44
	Total	186	4899	5085

Chi-squared test of independence, NS = Not significant, * $p < 0.05$, *** $p < 0.001$

We next examined the relationship between gene duplication and differential gene expression across the sexes (worker vs drone) (Table 4-2). We found that the proportion of sex-biased genes differed significantly between duplicates and singletons for both datasets focused on sex differences (Ashby *et al.*: $\chi^2_{df=1} = 30.78$, $p < 0.0001$; Vleurinck *et al.*: $\chi^2_{df=1} = 12.1$, $p = 0.0005$, χ^2 test of independence). Both analyses showed a greater frequency of sex-biased, and associated lower frequency of unbiased, genes among the duplicates

than the singletons (Table 4-2). We found that very few genes were differentially expressed between nurse and forager worker behavioral types (Table 4-3). Therefore, we did not perform similar tests between nurses and foragers. Regardless, overall, we found that genes showing biased expression tended to be more common than expected among duplicated genes than singletons when considering caste and sex differences.

Table 4-2 Observed (and expected) counts of genes differentially expressed across sexes for duplicated genes and singletons in two RNAseq datasets comparing worker and drone gene expression differences in A. mellifera.

Dataset	Expression	Duplicates	Singletons	Total
Ashby *et al.*	Drone-biased	25 (8.62)	200 (216.38)	225
***	Unbiased	167 (188.25)	4744 (4722.75)	4911
	Worker-biased	12 (7.13)	174 (178.87)	186
	Total	204	5118	5322
Vleurinck *et al*.	Drone-biased	4 (0.79)	17 (20.21)	21
***	Unbiased	191 (197.19)	5064 (5057.81)	5255
	Worker-biased	5 (2.03)	49 (51.97)	54
	Total	200	5130	5330

Chi-square test of independence, *** $p < 0.001$

Table 4-3 Observed (and expected) counts of genes differentially expressed between A. mellifera worker behavioral phenotypes for duplicated genes and singletons in an RNAseq dataset comparing nurse and forager gene expression differences in A. mellifera.

Dataset	Expression	Duplicates	Singletons	Total
Jasper *et al.*	Forager-biased	1 (0.04)	0 (0.96)	1
	Unbiased	214 (215.88)	5178 (5176.12)	5392
	Nurse-biased	1 (0.08)	1 (1.92)	2
	Total	216	5179	5395

Next, we compared level of caste-biased expression (as opposed to the number of caste-biased genes) between duplicate genes and singletons (Figure 4-2A-C). In this case, we found significant differences in the level of caste-biased expression between duplicates and singletons in all three studies that examined caste differences (Ashby et al.: W = 5.8e+05, p = 0.0047 Vleurinck et al.: W = 6.7e+05, p < 0.0001; Cameron et al.: W = 5.3e+05, p < 0.0001, Wilcoxon rank-sum test). In particular, duplicated genes tended to display significantly higher levels of caste-biased expression. In addition, duplicates had a higher level of sex-biased expression compared to singletons in the two datasets examined (Vleurinck et al.: W = 6.4 e+05, p < 0.0001; Ashby et al.: W = 6.4e+05, p < 0.0001, Wilcoxon rank-sum test) (Figure 4-2D-E). Duplicates also had a higher level of differential expression in comparisons between nurses and foragers (W = 6.9e+05, p < 0.0001, Wilcoxon rank-sum test) (Figure 4-2F). Finally, duplicates displayed a substantially and

significantly higher level of tissue-biased expression than singletons (W = 7.0 e+05, p < 0.0001, Wilcoxon rank-sum test) (Figure 4-2G).

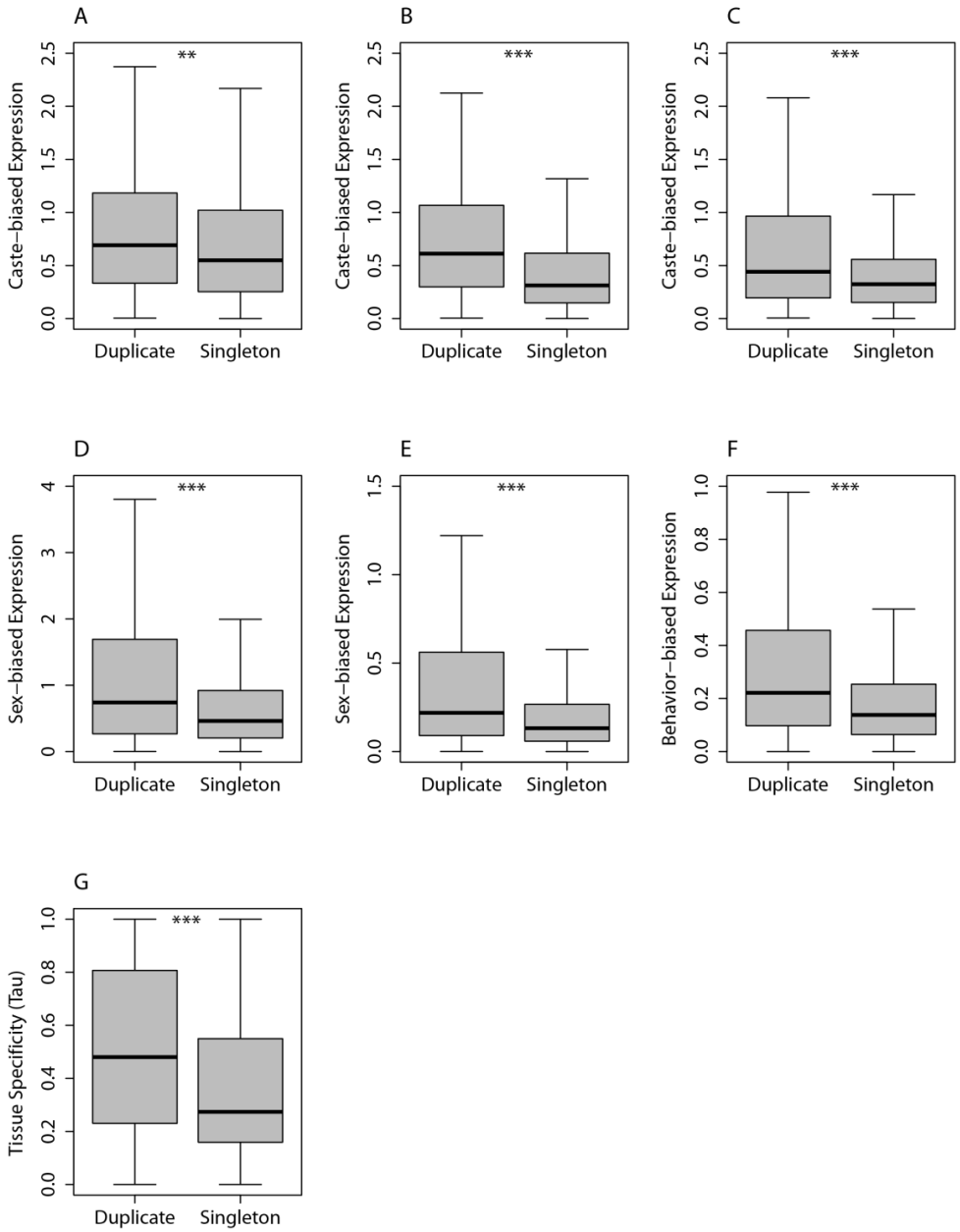

Figure 4-2 Biased gene expression calculated as the absolute value of the log2-fold change in expression for duplicated genes and singletons. Caste-biased expression from (A) Ashby et al., (B) Vleurinck et al., and (C) Cameron et al. Sex-biased expression from (D) Ashby et al. and (E) Vleurinck et al. Behavior-bias expression

from (F) Jasper et al. (G) Tissue-biased expression (Tau) from Jasper et al. Error bars represent standard errors. ** p < 0.01;* p < 0.001.**

We investigated the correlations of expression-bias within pairs of duplicate genes. Specifically, we were interested in determining if a gene that showed relatively high caste-biased expression, for example, also displayed high levels of sex-biased, behavior-biased, and tissue-biased expression. We first investigated the correlation of caste-biased expression using all genes found in the analyses of Ashby *et al.*, Vleurinck *et al.*, and Cameron *et al.*. We found that the correlations ranged from 0.200 to 0.266 (all pairwise comparisons $p < 0.0001$). In addition, the correlation between Ashby *et al.* and Vleurinck *et al.* sex-biased expression was 0.285 ($p < 0.0001$). Thus there is substantial evidence that genes that show biased expression in one type of analyses tend to show bias in others.

In order to determine the prevalence of such correlations within duplicated genes, we considered the partial correlation matrix derived from a MANOVA (Table 4-4). We found that most of the partial correlations were positive, indicating that there were associations in expression bias for duplicate gene pairs. However, there were two comparisons that resulted in a negative correlation. Nevertheless, as a whole, the partial correlations did indicate that there was a relationship between expression bias for paralogs, revealing that a paralog that showed substantial expression bias in one phenotypic context was likely to show substantial expression bias in another.

Table 4-4 Partial correlation matrix between measures of biased expression within duplicate gene pairs.

	A_Caste	J_Tissue	J_Behavior	V_Sex	V_Caste	C_Caste
A_Sex	0.139	0.218	0.083	0.21	0.117	0.156
A_Caste		0.183	0.017	0.029	0.188	-0.127
J_Tissue			0.216	0.204	0.190	-0.089
J_Behavior				0.064	0.159	0.098
V_Sex					0.484	0.091
V_Caste						0.18

A = Ashby *et al.*, V =Vleurinck *et al.*, C = Cameron *et al.*, J = Jasper *et al.*

4.4.3 Gene expression correlation between duplicate pairs

We compared expression classes of duplicate pairs in order to determine if the proportion of pairs showing discordant expression between phenotypes differed from random expectations. We found that a majority of the duplicate pairs displayed concordant caste-, sex-, and behavior-biased expression patterns (Table 4-5). We created a null distribution of pairs in order to test for the overrepresentation of certain expression pair classes. We did not find significant differences in the observed and expected expression pair classes for castes (Cameron *et al.*: $\chi^2_{df=1} = 0.06$, $p = 0.807$; Vleurinck *et al.*: $\chi^2_{df=1} = 2.2$, $p = 0.138$; Ashby *et al.*: $\chi^2_{df=1} = 0.6$, $p = 0.4396$, χ^2 test) (Table 4-5). In contrast, when

we compared the expression pair classes for sex-biased genes to the null distribution, we saw a significant difference between observed and expected classes for one out of the two datasets (Ashby *et al.*: $\chi^2_{df=1} = 6.11$, $p = 0.0134$; Vleurinck *et al.*: $\chi^2_{df=1} = 0.06$, $p = 0.8051$). Overall, however, paired expression classes were generally found at the frequency expected for the datasets.

Table 4-5 Observed and expected numbers of pairs of caste-, sex-, and behavior-biased gene expression classes showing correlations of expression classes among duplicate genes. 'Concordant' indicates that the duplicates had the same direction of expression bias (e.g., were both queen-biased) whereas 'Discordant' indicates that the duplicate genes showed different expression patterns of expression bias (e.g., one was queen-biased and the other worker-biased).

Phenotype	Dataset	Expression	Observed	Expected
Caste	Ashby *et al.*	Concordant	79	74.32
		Discordant	11	15.68
		Total	90	90
	Vleurinck *et al.*	Concordant	76	67.8
		Discordant	10	18.2
		Total	86	86
	Cameron *et al.*	Concordant	74	72.19
		Discordant	5	6.81
		Total	79	79
Sex	Ashby *et al.* *	Concordant	76	60.83
		Discordant	14	29.17
		Total	90	90
	Vleurinck *et al.*	Concordant	81	79.18
		Discordant	5	6.82
		Total	86	86
Behavior	Jasper *et al.*	Concordant	98	98.02
		Discordant	2	1.99
		Total	100	100.01

Chi-squared test, * $p < 0.05$

We next investigated if expression divergence between paralogs depended on relative location of genes in the genome (Figure 4-3). We found that paralogs on different linkage groups had similar levels of expression divergence to those on the same linkage group (Caste: Ashby *et al.*: W = 947, p = 0.8761; Vleurinck *et al.*: W = 894, p = 0.7921; Cameron *et al.*: W = 879, p = 0.1479; Sex: Ashby *et al.*: W = 972, p = 0.7139; Vleurinck *et al.*: W = 927, p = 0.5766; Behavior: Jasper *et al.*: W = 0.5627, p = 0.5627, Wilcoxon rank-sum test). There were also no significant differences in the level of tau, which defines tissue-specific expression, between duplicates on the same or different linkage groups (W = 1428, p = 0.0738) (Figure 4-3G).

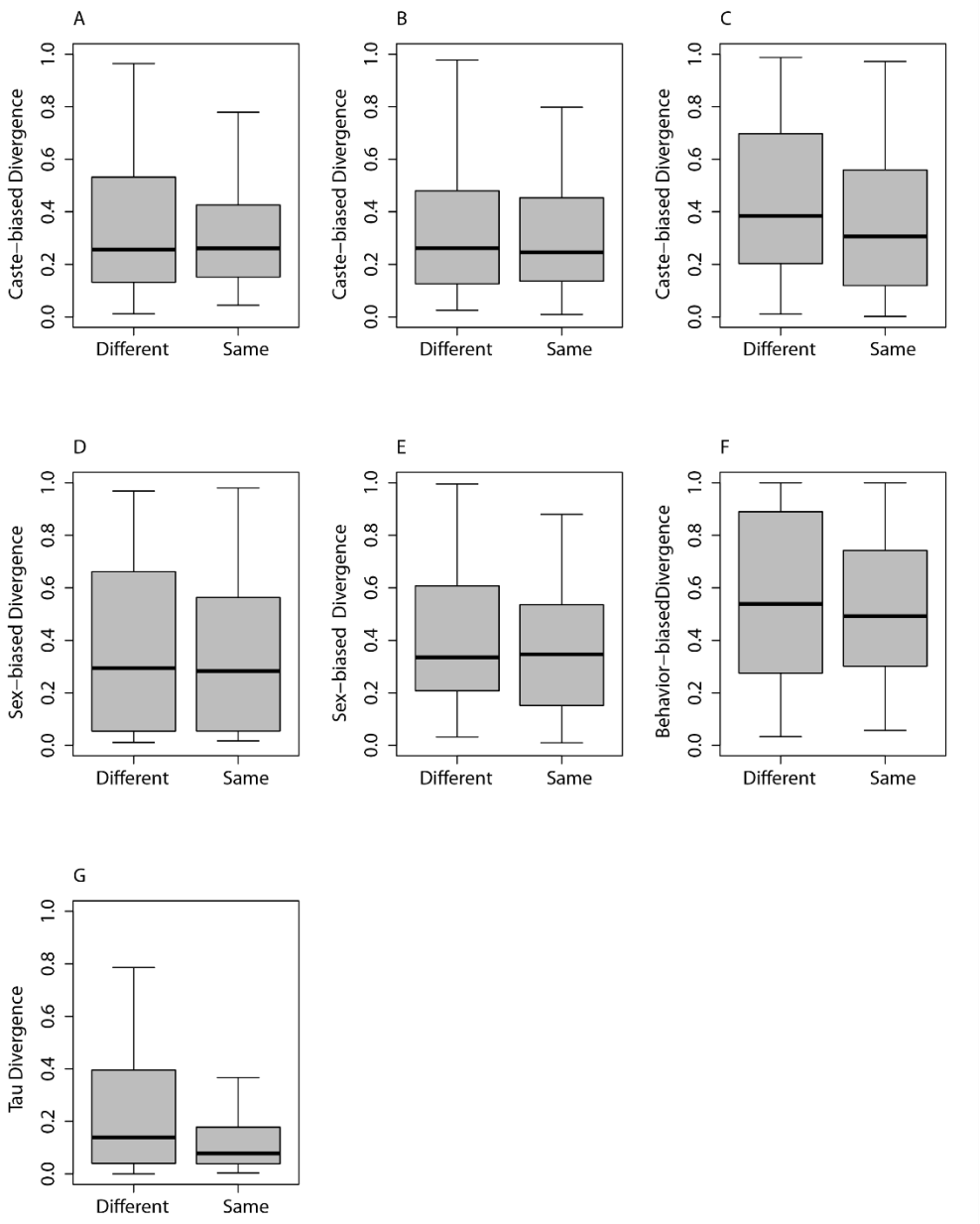

Figure 4-3 Divergence in biased gene expression between duplicate pairs on the same or different linkage group. Comparisons between castes from (A) Ashby et al. (B) Vleurinck et al. and (C) Cameron et al., between sexes from (D) Ashby et al. (E) Vleurinck et al., between worker behavioral types from (F) Jasper et al., and among tissues (Tau) from (G) Jasper et al.

4.4.4 Classification of evolutionary processes maintaining duplicate genes

We investigated the evolutionary processes maintaining duplicate genes in *A. mellifera* (Assis and Bachtrog 2013, 2015). We found that there were 63 cases of conservation, 28 cases of neofunctionalization (fifteen of D1 copy, the duplicate with higher sequence similarity to the single copy ortholog, thirteen of D2 copy, the duplicate with lower sequence similarity to the single copy ortholog), 9 cases of specialization, and no cases of subfunctionalization.

We next investigated evolutionary constraint (dN/dS) and relative expression across alternative phenotypes for genes that arose through conservation. We did not find a significant difference in the level of dN/dS between the duplicate pairs that were subject to conservation (W = 1.3+e03, p = 0.4227, Wilcoxon rank-sum test) (Figure 4-4A). Since conservation leads to duplicates maintaining similar functions, we expected similar levels of biased gene expression across conserved genes. There was no significant difference in the level of caste-biased expression ($\chi^2_{df=2} = 4.27$, $p = 0.118$, Kruskal-Wallis test) (Figure 4E) and sex-biased expression ($\chi^2_{df=2} = 0.34$, $p = 0.8414$) (Figure 4I) between single copy orthologs and the conserved duplicates.

The level of dN/dS was not significantly different between D1 and D2 for those duplicates maintained through specialization (W = 17, p = 0.2159, Wilcoxon rank-sum test) (Figure 4B). Duplicates that have undergone specialization are expected to have different levels of biased expression for the single copy ortholog and both duplicates. However, we found no difference in the level of caste- ($\chi^2_{df=2} = 3.58$, $p = 0.1671$, Kruskal-

Wallis test) or sex-biased gene expression ($\chi^2_{df=2} = 0.79$, $p = 0.6723$) between D1, D2, and single copy orthologs (Figure 4-4F, 4J).

Next, we examined the differences in dN/dS between duplicates that have undergone neofunctionalization. For those duplicates that underwent neofunctionalization of the D1 gene, there was a significantly higher level of dN/dS for the D1 copy (W = 29.5, $p = 0.036$, Wilcoxon rank-sum test) (Figure 4-4C). However, this was not the case for those duplicates in which D2 underwent neofunctionalization (W = 57, $p = 0.31$, Wilcoxon rank-sum test) (Figure 4-5D). We found no difference in caste-biased expression between the single copy ortholog and both duplicates (D1: $\chi^2_{df=2} = 0.52$, $p = 0.7705$, D2: $\chi^2_{df=2} = 2.31$, $p = 0.3142$, Kruskal-Wallis test) (Figure 4-4G, 4H). There was also no difference in the level of sex-biased expression between genes that underwent neofunctionalization of D1 ($\chi^2_{df=2} = 1.78$, $p = 0.4115$, Kruskal-Wallis test) (Figure 4-4K). Though, we saw that D2 had a higher level of sex-biased expression compared the single copy ortholog and D1 copy for those duplicates that underwent neofunctionalization of D2 ($\chi^2_{df=2} = 7.1$, $p = 0.02876$, Kruskal-Wallis test) (Figure 4-4L).

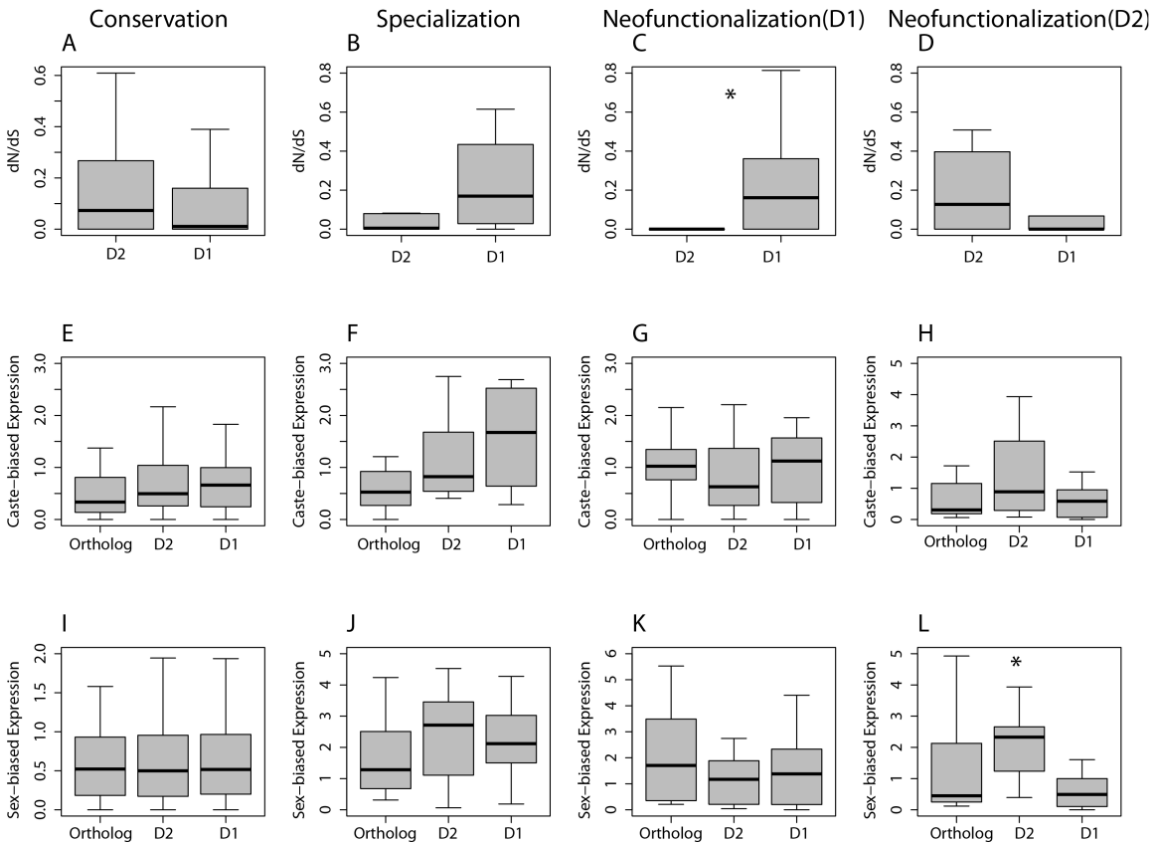

Figure 4-4 Comparison of metrics for duplicated genes maintained through conservation, neofunctionalization (D1 copy and D2 copy), and specialization. D1 and D2 are the *A. mellifera* duplicate genes with higher and lower sequence similarity to the single copy *B. terrestris* ortholog, respectively. Error bars represent standard errors. (A-D) Mean levels of dN/dS for duplicate pairs. (E-H) Caste-biased expression, as measured by absolute value of the $\log_2$ fold change in expression between queens and workers. (I-L) Sex-biased expression, as measured by absolute value of the $\log_2$ fold change in expression between drones and workers.

Expression patterns of single copy orthologs might limit the evolutionary processes maintaining a duplicate pair in the genome. Therefore, we examined the level of differential expression of single copy orthologs in *B. terrestris* of *A. mellifera* gene duplicates to gain insight into possible constraints on expression evolution of duplicated genes (Figure 5). We found a significant difference in the level of sex-biased expression between single copy orthologs in *B. terrestris* that have undergone specialization, neofunctionalization, and

conservation, with orthologs that underwent neofunctionalization of the D1 copy in *A. mellifera* having the highest level ($\chi^2_{df=3} = 9.18$, $p = 0.027$, Kruskal-Wallis test) (Figure 4-5). However, this trend was not found for genes displaying caste-biased expression ($\chi^2_{df=3} = 6.48$, $p = 0.9039$, Kruskal-Wallis test).

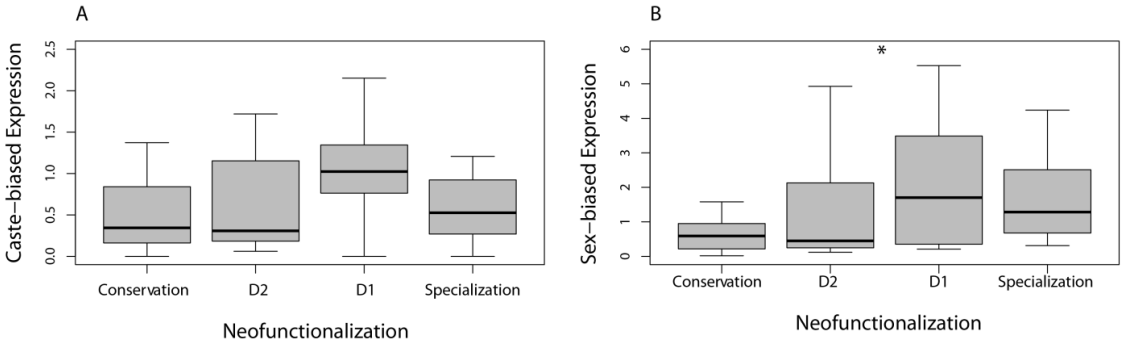

Figure 4-5 Comparison of biased expression of single copy orthologs in B. terrestris that have been duplicated in A. mellifera and been maintained through different evolutionary processes. (A) Levels of caste-biased expression of single copy orthologs in B. terrestris (χ2 df = 3 = 6.4817, p = 0.09039, Kruskal-Wallis test). (B) Sex-biased expression between males and workers (female) (χ2 df = 3 = 9.1811, P = 0.02698, Kruskal-Wallis test). * p < 0.05.

4.5 Discussion

4.5.1 Rates of species-specific gene duplication

Sociality has arisen multiple times in insects. This phenomenon has been of great interest to researchers and has been the focus of many genomic studies aimed at examining the genetic changes associated with the evolution of sociality (Woodard et al. 2011; Harpur et al. 2014; Roux et al. 2014; Kapheim et al. 2015). We were interested in the hypothesis that gene duplication has facilitated the evolution of sociality and caste differences in insect societies. We observed a positive correlation between social complexity and the rate of

species-specific gene duplication. This suggests that more highly social bee taxa possess higher rates of gene duplication or lower rates of duplicate gene loss. However, this correlation was not significant with phylogenetic correction. Regardless, the number of species examined in this study was modest and the strength of the correlation was substantial. Therefore, further investigation is needed to determine whether our observation between gene duplication and sociality plays a role in the evolution of complex societies.

4.5.2 Duplicated genes and biased gene expression

We hypothesized that gene duplication provided new copies of genes which could be co-opted into the development of social insect phenotypes. We thus expected an enrichment of caste-biased genes in duplicates compared to singletons. We did, in fact, find significantly more caste-biased and sex-biased genes in duplicated genes when compared to singletons (Table 4-1;Table 4-2).

Our findings of significant excesses of phenotype-biased genes among duplicates agree with previous studies performed in *D. melanogaster* and *C. elegans* sexes (Cutter and Ward 2005; Wyman et al. 2012). These prior studies found enrichment for duplicates showing phenotype-biased expression. In addition, these investigations uncovered an excess of duplicates with male-biased gene expression, suggesting that gene duplication is frequently involved in the evolution of male-biased traits (Cutter and Ward 2005; Wyman et al. 2012). In our comparison of sex-biased gene expression between female workers and male drones, we found more drone- and worker-biased duplicates than expected. The datasets we used included individuals from the larval and pupal stages. So it is possible that any strong male effects would not have been identified because we may not have

detected the full array of differential gene expression found in adults (Morandin et al. 2015; Ashby et al. 2016; Lockett et al. 2016; Vleurinck et al. 2016).

We expected that duplicates would display a higher level of biased expression compared to singletons. Such a finding would be consistent with the hypothesis that duplicate gene expression can be co-opted into the evolution of different phenotypic forms (Gadagkar 1997a; Gallach and Betran 2011). We did find that duplicates tended to have higher levels of caste-, sex-, behavior-, and tissue-biased expression compared to singletons (Figure 4-2). Overall, these results agree with past studies that found that duplicates tended to become more specialized in their expression patterns (Freilich et al. 2006; Farré and Albà 2010; Assis and Bachtrog 2013). Similarly, gene families of increasing size have been found to show increasing levels of expression bias (Huminiecki and Wolfe 2004; Kapheim et al. 2015). Our results do suggest that gene duplication does permit for the evolution of variation in expression levels and may allow for phenotypic diversification at multiple phenotypic levels.

We further examined the expression patterns of pairs of duplicate genes to determine if they showed concordant expression patterns between different castes, sexes, and worker behavioral types. In general, we did not find significant enrichment of duplicate pairs with concordant expression relative to expectations (Table 4-5). However, we did see enrichment for duplicates with the similar expression bias for the Ashby et al. dataset when comparing sex-biased expression (Table 4-5). This excess of duplicate pairs with concordant expression was observed in analysis of Drosophila sexes (Wyman et al. 2012). Thus it appears that duplicate gene pairs may maintain similar expression profiles to each other. This might reflect the fact that a new duplicate gene is likely to have maintained its

expression profile and function immediately after duplication, and that it takes time for a discordant expression profile to evolve. For example, there is some evidence that duplication of a gene that is already sex-biased may allow the gene's paralog to become even more sex-biased (Wyman et al. 2012). The result that duplicates tend to have higher levels of biased gene expression but tend not to differ in their directional bias may be indicative of this mechanism.

We found that paralogs located on the same chromosome did not necessarily have similar expression patterns compared to paralogs located on different chromosomes (Figure 4-3). This result differs from previous studies (Mikhaylova et al. 2008) and suggests that genes on the same chromosome are not necessarily subject to similar regulatory regimes (Ibn-Salem et al. 2016; Lan and Pritchard 2016). Therefore, a new gene duplicate may evolve divergent expression patterns from its parent paralog, even if it is duplicated onto the same chromosome.

We examined the correlation between caste-, sex-, behavior-, and tissue-biased expression for individual duplicate pairs (Table 4-4). A majority of correlations between these different measures of phenotype-biased expression were positive. This indicates that duplicates that are differentially expressed in one phenotypic context tend to be differentially expressed in other contexts (Hunt et al. 2013). Genes with higher levels of differential expression may be subjected to weakened selective constraint on gene expression compared to genes that are more uniformly expressed among phenotypes (Mank and Ellegren 2009; Hunt et al. 2011; Leichty et al. 2012). Therefore, loci experiencing weak selective constraint may be more likely to be differentially expressed in a variety of contexts (Hunt et al. 2011; Leichty et al. 2012).

Our results suggest that gene duplication may provide genetic material that can be co-opted in the evolution of alternative phenotypes. However, there are other mechanisms that can potentially explain the patterns that we observed. For example, it is possible that ancestral genes that were already differentially expressed between phenotypes were more likely to duplicate because of mutation bias. In addition, copy number variants of differentially expressed genes could be less likely to be under purifying selection, leading to fixation of such genes (Cardoso-Moreira et al. 2016). Or, the genome may be more tolerant of the acquisition of phenotype-biased genes compared to singletons, particularly if phenotypic-biased genes are not essential (Mank and Ellegren 2009). Therefore, biased duplicates may be fixed at a higher rate than biased singletons. Thus there are potentially several molecular evolutionary mechanisms that could lead to the observation of a correlation between phenotype-biased expression and gene duplication.

4.5.3 Evolutionary Processes maintaining duplicates

We investigated the processes that maintained duplicate genes within the *A. mellifera* genome. This analysis uses gene expression as a proxy for gene function. Therefore, we must keep in mind that there are possibilities for gene duplicates to diverge in function but not differ in their expression pattern. Regardless, we observed that conservation, neofunctionalization, and specialization were the primary evolutionary processes associated with gene duplication in A. mellifera. Interestingly, we identified no cases of subfunctionalization.

Conservation was found to be the most common process maintaining gene duplicates in *A. mellifera*. It is notable that prior studies also found that conservation was

one of the most common mechanisms maintaining gene duplicates in mammals and plants (Assis and Bachtrog 2015; Wang et al. 2016). In contrast, neofunctionalization was found to be the most common process maintaining gene duplicates in Drosophila (Assis and Bachtrog 2013). The difference between these findings could be due to the differences in effective population size among the studied taxa. Effective population size is predicted to be correlated with efficacy of natural selection. Natural selection is less efficient in smaller populations. Drosophila, with its large effective population size, may have more neofunctionalized genes maintained by selection. In contrast, natural selection will operate less efficiently in species with smaller effective population size, such as A. mellifera and mammals, allowing potentially neofunctionalized genes to be fixed less often (Jensen and Bachtrog 2011; Romiguier et al. 2014; Galtier 2016).

Interestingly, there was a notable lack of subfunctionalization across all studied taxa (Assis and Bachtrog 2015; Lan and Pritchard 2016). This is surprising because it has been suggested that subfunctionalization is an important process in the retention of duplicate genes (Lynch and Conery 2000). Subfunctionalization requires that both duplicates start off with the same function and are in dosage balance. Therefore, subfunctionalization is more likely to occur for large scale duplications like whole genome duplication events, which maintain the regulatory environments of the focal genes (Casneuf et al. 2006; Fares et al. 2013). Indeed, past studies have discovered a greater likelihood of subfunctionalization after whole genome duplication (Fares et al. 2013). The lack of observed subfunctionalization in our analyses could also be due to the datasets used for classification. The analysis classifying duplicates into evolutionary processes was performed using an expression profile across whole-body A. mellifera and B. terrestris

queens, workers, and drones. This may lead to an underestimation of potential expression differences across tissues and time, which may obfuscate some patterns of subfunctionalization (Assis and Bachtrog 2013, 2015).

We examined differences in the level of caste-biased expression of single copy orthologs for duplicate genes maintained by conservation, neofunctionalization, and specialization. Duplicates that underwent conservation tended to arise from single copy orthologs that had lower levels of differential expression (Figure 4-5). The low level of differential expression suggests that duplicates that have undergone conservation are more essential and broadly expressed than those that have undergone neofunctionalization and specialization. Genes that are subject to the latter mechanisms generally displayed biased expression among phenotypes, leading to the development of new functions. This suggests that the ancestral function of a pair of duplicates may limit their evolutionary trajectory (Wang et al. 2016). We also examined evolutionary and expression characteristics of duplicates that were maintained through the different evolutionary processes (Assis and Bachtrog 2013, 2015). We found no significant difference in the constraint (dN/dS) between duplicate pairs involved in conservation (Figure 4-4A). However, we identified differences in dN/dS between duplicates that underwent neofunctionalization of the D1 copy (Figure 4-4C). This is interesting given that the duplicate that gains the new function, D1, has a higher rate of dN/dS.

4.5.4 Conclusions

Recently, considerable attention has been paid to the role of novel genes in the evolution of phenotypic diversity in social species (Johnson and Tsutsui 2011; Tautz and

Domazet-Lošo 2011; Feldmeyer et al. 2014; Sumner 2014; Jasper et al. 2015). This study provides further insight on the role of new genes, created through the process of gene duplication, in the evolution of insect societies. More highly social bee species may have higher gene duplication rates. Duplicate genes seem to be preferentially co-opted into caste- and sex-specific function. Moreover, duplicated genes are apparently subject to conservation, neofunctionalization, and specialization in *A. mellifera*. Overall, this study adds to the accumulating evidence that gene duplication has played a substantial role in the evolution of complex societies, in particular, and alternative phenotypes, in general.

CHAPTER 5. CONCLUSIONS

This thesis consists of three studies aimed at furthering our understanding of eusociality across a spectrum of social species. Chapter two examines the population genetics of an invasive wasp in its native and invasive range. Chapter three investigates the population genetics and sex ratio of captive naked mole rat populations. Chapter four focuses on gene duplication and the role it plays in the evolution of the caste system in the honey bee, *A. mellifera*.

Social insects make up a large percent of invasive species. This enrichment is thought to be due to their social structure; therefore, it is important to study the population genetics of invasive social animals and other factors that may lead to invasion success. Chapter two, *Population genetic structure of the predatory, social wasp Vespula pensylvanica in its native and invasive range*, examined *V. pensylvanica's* invasion of the Hawaiian Islands. This invasion is particularly interesting because the geography of ocean archipelagos may pose a challenge to the invasion dynamics of a eusocial species. We saw that there was little genetic isolation in *V. pensylvanica's* native range which spans over 2000 km in North America, suggesting that these wasps can disperse over large ranges. Also, there was the presence of substantial genetic structure across *V. pensylvanica's* invasive range, which may reflect the role of geographic barriers on gene flow in invasive species. This drastic difference in genetic structure between the native and invasive range would be interesting to follow up on, particularly the influence of human transportation on forming these patterns. Studies in other insects have tracked spread of invasive species along human modes of transportation like highways (Fonzi et al. 2015; Egizi et al. 2016).

Therefore it's important to understand the human factors associated with the spread of *V pensylvanica* to prevent further increase in its invasive range. Overall, this chapter provides insight on the invasion success of social insects and the constraint geography can have on shaping the genetic structure of invasive eusocial animals.

Sociality is highly concentrated within the Hymenoptera; however, naked mole rats are one of the rare eusocial mammals. Not many studies have been performed to examine the population structure and genetic diversity of this species. Chapter three, *Genetic diversity and differentiation of naked mole rats, Heterocephalus glaber, in zoo populations*, is the first study to examine the genetic diversity and structure of captive naked mole rats. In this study, we created a set of microsatellite primers used to study the levels of genetic diversity across populations in three zoos. I saw modest levels of diversity in the zoo populations, which supports the hypothesis that naked mole rat colonies are not as inbred as previously suspected. Early genetic studies suggested that high levels of inbreeding promoted the evolution of eusocial in naked mole rats (Reeve et al. 1990; Lacey and Sherman 1991; Faulkes et al. 1997a). This chapter adds to the growing literature which refutes this initial hypothesis and demonstrates that inbreeding is not required for the evolution of cooperation in mammals (Braude 2000; Ciszek 2000; Burland et al. 2002; Pemberton 2004; Ingram et al. 2015). Given this challenge of the inbreeding hypothesis for the evolution of eusociality in the naked mole rat, more hypotheses should be explored, such as the aridity-food distribution hypothesis (Jarvis et al. 1994; Faulkes and Bennett 2001). The aridity food distribution hypothesis states that eusociality evolved in naked mole rats in response to ecological constraints, like unpredictable rainfall, that limit food distribution, dispersal and new colony formation (Faulkes et al. 1997b). There has been

some evidence for the association between social group size and rainfall variation across seven different mole rat species, which vary in level of sociality, so it would interesting to further examine this correlation between ecological constraints and level of sociality (Le Comber et al. 2002). We also saw some evidence for population genetic structure across the zoo populations, which may reflect the founding and husbandry of these populations. There was no evidence for sex ratio bias in any of the zoo populations. Sex ratio adjustment has been linked to the evolution of eusociality in haplodiploid social insects. Naked mole rats are diploid; therefore, inclusive fitness predicts an equal sex ratio for this species. An equal sex ratio also shows a lack of condition-dependent sex ratio adjustment, so both males and females cost equal for the reproductive female to produce (Trivers and Willard 1973). This study gives us insight into the population genetics and sex ratio of a eusocial, diploid mammal.

Much research concerning eusociality is interested in the development and evolution of castes. Chapter four, *Gene duplication and the evolution of phenotypic diversity in insect societies*, suggests that gene duplication may have aided the evolution of castes and sexes. In 1997, Gadagkar (Gadagkar 1997b) proposed the idea of "genetic release", in which a gene duplication event or significant variation in gene expression could free genes from the effects of stabilizing selection and allow social organisms to evolve castes (Gadagkar 1997). Therefore, I examined this hypothesis that gene duplication creates genetic material that can be co-opted in the evolution of castes, utilizing modern genomic techniques. I saw a correlation between duplication rate and level of sociality across ten bees in the *Apoidea*. Duplicate genes also had a higher level of differential expression when compared to single-copy genes. Furthermore, I saw that there was an

excess of duplicate genes with biased expression. Also, most duplicate genes were maintained in the *A. mellifera* genome by conservation, which was also found in mammals and plants (Assis and Bachtrog 2015; Wang et al. 2016). Therefore, I found a potential role for gene duplication in the evolution of eusociality. To follow up on this study, we can examine this association between biased gene expression and gene duplication across species with varying levels of sociality. For example, we can look at differential expression and gene duplication in a solitary bee species like *Megachile rotundata* or a species that displays simple eusociality like *Bombus terrestris*. Under the hypothesis that gene duplication aids the evolution of alternative phenotypes, we would see an increased level of caste biased gene expression in social bee species and increased levels of sex biased gene expression across sexes amongst duplicates. Also, it would be interesting to examine different aspects of gene regulation, such as DNA methylation and *cis*-regulatory sequences that may have led to this association between gene expression and gene duplication.

Eusociality is a rare and complex form of sociality that has evolved multiple times across the tree of life (Rehan et al. 2012). However, most studies have focused on the social insects. My research attempts to examine multiple aspects of the evolution of eusociality in social insects and eusocial mammals. I explored the population genetics structure of both invasive and captive social organisms. Also, I examined the role of gene duplication in the evolution of castes. With the increased availability of sequencing and advancing technology, this larger phylogenetic perspective towards eusociality may likely aid in identifying common trends related to the evolution of eusociality.

APPENDIX A. SUPPLEMENT MATERIAL FOR CHAPTER 2

Table A 1 Locations and total numbers of *V.pensylvanica* wasps collected from traps in the sampled transects, ranges, and regions (NA = location not determined).

Range	Region	Transect	Trap	Number	Latitude	Longitude
Native	California	Atascandero	AT1	5	-120.6625105	35.49163266
Native	California	Atascandero	AT2	5	-120.6488526	35.50535156
Native	California	Atascandero	AT3	5	-120.6411037	35.49088159
Native	California	Atascandero	AT4	5	-120.6282105	35.48072472
Native	California	Atascandero	AT5	5	-120.6234256	35.46245113
Native	California	Atascandero	AT6	5	-120.6326374	35.44985656
Native	California	Atascandero	AT7	5	-120.6177986	35.4342955
Native	California	Atascandero	AT8	4	-120.6043978	35.4194982
Native	California	Balboa Park	BP1	5	NA	NA
Native	California	Balboa Park	BP2	1	-117.1510313	32.72618921
Native	California	Balboa Park	BP3	5	NA	NA
Native	California	Balboa Park	BP4	5	NA	NA
Native	California	Balboa Park	BP5	2	NA	NA
Native	California	BR	BR1	5	NA	NA
Native	California	BR	BR2	2	-120.7819324	36.16771691
Native	California	BR	BR3	5	-120.7575863	36.14742337
Native	California	BR	BR4	5	-120.7438367	36.13030809
Native	California	BR	BR5	2	-120.7278427	36.11650741
Native	California	BR	BR6	5	-120.6907588	36.09930088
Native	California	BR	BR7	2	-120.6702896	36.08988321
Native	California	BR	BR8	3	-120.6519382	36.07099627
Native	California	Corning	CN3	2	NA	NA
Native	California	Corning	CN5	5	NA	NA
Native	California	Corning	CN6	1	NA	NA
Native	California	Diablo	DB1	5	-121.9882869	37.89821437
Native	California	Diablo	DB2	5	-121.9835124	37.89120799
Native	California	Diablo	DB3	5	-121.9753211	37.88347235
Native	California	Diablo	DB4	5	-121.9649215	37.88018497
Native	California	Diablo	DB5	5	-121.9601611	37.8733413
Native	California	Diablo	DB6	2	-121.9499387	37.87224066
Native	California	Diablo	DB7	5	-121.9438767	37.86612278
Native	California	Diablo	DB8	5	-121.940417	37.87131976
Native	California	Diablo	DB9	5	NA	NA
Native	California	Diablo	DB10	1	-122.0088063	37.91647466
Native	California	La Jolla	LJ2	5	-117.245734	32.87161157
Native	California	La Jolla	LJ3	4	-117.244134	32.88785367

Table A 1 continued

Range	Region	Transect	Trap	Number	Latitude	Longitude
Native	California	Lake Shasta	SL2	2	-122.3981411	40.64829468
Native	California	Lake Shasta	SL3	2	-122.3970916	40.66606955
Native	California	Lake Shasta	SL4	5	-122.4034735	40.68368485
Native	California	Lake Shasta	SL5	4	-122.3939879	40.69000328
Native	California	Lake Shasta	SL6	5	-122.3758071	40.68211463
Native	California	Lake Shasta	SL7	5	-122.3198486	40.80545289
Native	California	Lake Shasta	SL8	4	-122.3107489	40.80327509
Native	California	Los Padres	LP1	4	NA	NA
Native	California	Los Padres	LP2	5	-119.3718463	34.62534394
Native	California	Los Padres	LP3	4	-119.2831089	34.51472353
Native	California	Los Padres	LP4	5	-119.2796706	34.50547854
Native	California	Los Padres	LP5	5	-119.3006382	34.50034458
Native	California	Los Padres	LP6	5	-119.2973696	34.4845709
Native	California	Los Padres	LP7	5	-119.2874708	34.46910532
Native	California	Los Padres	LP8	5	-119.2752783	34.45235339
Native	California	Los Padres	LP9	5	NA	NA
Native	California	Morgan Hill	MH1	5	-121.5969351	37.13665128
Native	California	Morgan Hill	MH2	5	-121.5924805	37.14693049
Native	California	Morgan Hill	MH4	5	-121.5881863	37.15518379
Native	California	Morgan Hill	MH5	5	-121.598386	37.16265905
Native	California	Portrero Road	PO1	1	NA	NA
Native	California	Portrero Road	PO2	3	-119.0119625	34.15735841
Native	California	Portrero Road	PO4	2	-118.9515783	34.15464988
Native	California	Portrero Road	PO5	2	-118.9347434	34.15693864
Native	California	Portrero Road	PO6	5	-118.9209383	34.15556296
Native	California	Portrero Road	PO7	5	-118.9038424	34.15256182
Native	California	Portrero Road	PO8	5	NA	NA
Native	California	Ramona	RM1	5	NA	NA
Native	California	Ramona	RM2	5	-117.0085586	33.09572365
Native	California	Ramona	RM3	5	-116.9937408	33.09096695
Native	California	Ramona	RM5	4	-116.9461612	33.09054209
Native	California	Ramona	RM6	1	-116.9327132	33.0856955
Native	California	Ramona	RM7	5	-116.9220089	33.08498111
Native	California	Ramona	RM8	5	-116.9029612	33.07438928
Native	California	Ramona	RM9	5	NA	NA
Native	California	Ramona	RM10	2	-116.9306462	33.01089996
Native	California	Ramona	RM11	5	-116.9520977	33.01264618
Native	California	Ramona	RM12	1	-116.9765623	32.99417325
Native	California	Santa Maria	SM1	4	NA	NA
Native	California	Santa Maria	SM2	5	NA	NA
Native	California	Santa Maria	SM3	5	-120.2947794	35.01744283
Native	California	Santa Maria	SM4	4	-120.2755065	35.0276189

Table A 1 continued

Range	Region	Transect	Trap	Number	Latitude	Longitude
Native	California	Santa Maria	SM5	5	-120.2553093	35.0268146
Native	California	Santa Maria	SM6	5	-120.2344669	35.02445978
Native	California	Santa Maria	SM7	1	-120.2163076	35.02071577
Native	California	Santa Maria	SM8	2	-120.192963	35.02947343
Native	California	Tilden Park	TL1	5	-122.262673	37.87251217
Native	California	Tilden Park	TL2	5	-122.2626753	37.87249976
Native	California	Tilden Park	TL3	5	-122.2645472	37.90857365
Native	California	Tilden Park	TL4	5	-122.260125	37.90371266
Native	California	Tilden Park	TL5	5	-122.2557873	37.89788828
Native	California	Tilden Park	TL7	4	-122.2422722	37.89360513
Native	California	Tilden Park	TL8	5	NA	NA
Native	California	Tilden Park	TL9	5	-122.2440586	37.90513899
Native	California	Tilden Park	TL10	5	-122.2351524	37.89956373
Native	California	Tilden Park	TL11	3	-122.2252455	37.89882909
Native	California	Tilden Park	TL12	5	-122.2181079	37.89876462
Native	California	Tilden Park	TL13	5	-122.2070842	37.90237529
Native	California	Tilden Park	TL15	1	-122.200961	37.91382539
Native	California	Tres Pinos	TP1	1	-121.3513815	36.8135695
Native	California	Tres Pinos	TP2	5	-121.3247985	36.79300312
Native	California	Tres Pinos	TP3	5	-121.3092239	36.77326415
Native	California	Tres Pinos	TP4	4	-121.2953895	36.75661871
Native	California	Tres Pinos	TP5	5	-121.2827306	36.74605407
Native	California	Tres Pinos	TP6	3	-121.2487435	36.66657502
Native	California	Tres Pinos	TP7	5	-121.2262774	36.63081232
Native	California	Tres Pinos	TP8	1	-121.2101488	36.61409636
Native	Colorado	Within Ft Collins	A	5	-105.1489584	40.58976202
Native	Colorado	Within Ft Collins	B	5	-105.1336695	40.5953513
Native	Colorado	Within Ft Collins	C	3	-105.0233845	40.53425222
Native	Colorado	Within Ft Collins	D	2	-105.133885	40.56067066
Native	Colorado	Within Ft Collins	E	5	-105.1013455	40.56805014
Native	Colorado	Within Ft Collins	F	5	-104.9968316	40.6100399
Native	Colorado	Within Ft Collins	G	5	-105.0805405	40.5989818
Native	Colorado	Within Ft Collins	H	5	-105.0719093	40.58555028
Native	Colorado	Outside Ft Collins	I	5	NA	NA
Native	Colorado	Outside Ft Collins	J	5	-104.9948452	40.52336638
Native	Colorado	Outside Ft Collins	K	5	-105.1394556	40.62114189
Native	Colorado	Outside Ft Collins	L	5	-105.0456954	40.57488405
Native	Colorado	Loveland	M	5	NA	NA
Native	Colorado	Loveland	N	5	-105.2240121	40.42633681
Native	Colorado	Outside Larimer County	Q	5	NA	NA
Native	Oregon	Chemult	CM1	2	NA	NA
Native	Oregon	Columbia River Gorge	CG7	1	-122.1993983	45.54533141

Table A 1 continued

Range	Region	Transect	Trap	Number	Latitude	Longitude
Native	Oregon	Columbia River Gorge	CG8	4	-122.2178433	45.53895201
Native	Oregon	Columbia River Gorge	CG9	2	NA	NA
Native	Oregon	Klamath Falls	KF1	4	NA	NA
Native	Oregon	Mill City	MC9	3	NA	NA
Native	Oregon	Mill City	MC11	1	NA	NA
Native	Oregon	Mill City	MC14	3	NA	NA
Native	Oregon	Mill City	MC15	1	NA	NA
Native	Oregon	Sisters	SS11	2	-121.6166661	44.36265427
Native	Oregon	Sisters	SS12	2	-121.6333834	44.37426218
Native	Oregon	Sisters	SS13	1	-121.6515137	44.37922303
Native	Oregon	Salem Area	SA11	5	-123.0638866	44.8578797
Native	Oregon	Salem Area	SA13	2	-123.0617018	44.8500788
Native	Oregon	Salem Area	SA14	5	-123.0617277	44.8421956
Native	Oregon	Salem Area	SA15	5	-123.0609565	44.83943
Native	Oregon	Salem Area	SA16	5	-123.0599322	44.8332039
Native	Oregon	Salem Area	SA18	5	-123.0586827	44.8207158
Native	Oregon	Salem Area	SA19	5	-123.0551936	44.8169159
Native	Oregon	Salem Area	SA110	3	-123.0498058	44.8122627
Native	Oregon	Salem Area	SA21	5	-123.0074599	44.8423212
Native	Oregon	Salem Area	SA22	1	-123.0075579	44.8381074
Native	Oregon	Salem Area	SA25	2	-123.0162291	44.8247436
Native	Oregon	Salem Area	SA27	2	-123.0238355	44.8170189
Native	Oregon	Salem Area	SA210	4	-123.0307534	44.8014567
Native	Oregon	Salem Area	SA31	4	-122.9577489	44.927899
Native	Oregon	Salem Area	SA32	1	-122.9510703	44.9279313
Native	Oregon	Salem Area	SA33	2	-122.9419943	44.9279268
Native	Oregon	Salem Area	SA34	2	-122.9345151	44.927334
Native	Oregon	Salem Area	SA36	5	-122.9152607	44.9271623
Native	Oregon	Salem Area	SA38	1	-122.8941081	44.9293407
Native	Oregon	Salem Area	SA39	5	-122.8832556	44.9282997
Native	Oregon	Salem Area	SA41	1	-122.9709487	45.1158094
Native	Oregon	Salem Area	SA42	4	-122.9810093	45.1169244
Native	Oregon	Salem Area	SA43	5	-122.988639	45.1249556
Native	Oregon	Salem Area	SA44	5	-123.0012631	45.130529
Native	Oregon	Salem Area	SA46	5	-123.0018524	45.1500796
Native	Oregon	Salem Area	SA47	2	-122.9977354	45.1596866
Native	Oregon	Salem Area	SA48	1	-122.9942007	45.1665795
Native	Oregon	Salem Area	SA410	3	-122.9862325	45.1859536
Native	Oregon	Salem Area	SA51	5	-122.8522715	45.1087737
Native	Oregon	Salem Area	SA52	2	-122.8564325	45.1009543
Native	Oregon	Salem Area	SA53	5	-122.8562574	45.0970499
Native	Oregon	Salem Area	SA54	3	-122.8586335	45.0933641

Table A 1 continued

Range	Region	Transect	Trap	Number	Latitude	Longitude
Native	Oregon	Salem Area	SA55	5	-122.8668189	45.0887533
Native	Oregon	Salem Area	SA56	5	-122.8666273	45.0840587
Native	Oregon	Salem Area	SA57	5	-122.8663722	45.0740713
Native	Oregon	Salem Area	SA58	1	-122.8728072	45.071954
Native	Oregon	Salem Area	SA510	5	-122.8797124	45.0720079
Native	Wyoming	Chugadul Caspar	R	4	-106.3631826	42.83038208
Native	Wyoming	Chugadul Caspar	S	1	-104.8218597	41.75640883
Native	New Mexico	Chimayo	T	5	-105.9470604	36.00088931
Invasive	Molokai	Forest Reserve Road	FR7	3	21.13263214	-157.0040807
Invasive	Molokai	Forest Reserve Road	FR8	5	21.13291126	-156.9955886
Invasive	Molokai	Forest Reserve Road	FR9	5	21.13650089	-156.988431
Invasive	Molokai	Forest Reserve Road	FR10	5	21.13998072	-156.9812962
Invasive	Molokai	Forest Reserve Road	FR11	5	21.13966816	-156.9733287
Invasive	Molokai	Forest Reserve Road	FR12	5	21.13939491	-156.9692314
Invasive	Molokai	Forest Reserve Road	FR13	4	21.1386392	-156.9651993
Invasive	Molokai	Forest Reserve Road	FR14	5	21.13825112	-156.9617134
Invasive	Molokai	Forest Reserve Road	FR15	5	21.13758543	-156.9576302
Invasive	Molokai	Forest Reserve Road	FR16	5	21.13893064	-156.9535437
Invasive	Molokai	Forest Reserve Road	FR17	5	21.13777771	-156.9496573
Invasive	Molokai	Forest Reserve Road	FR18	1	21.13710715	-156.945549
Invasive	Molokai	Forest Reserve Road	FR19	5	21.1349546	-156.9417994
Invasive	Molokai	Forest Reserve Road	FR20	5	21.13280732	-156.9385996
Invasive	Molokai	Forest Reserve Road	FR21	5	21.13300622	-156.9341812
Invasive	Molokai	Forest Reserve Road	FR22	5	21.13221858	-156.9297457
Invasive	Molokai	Forest Reserve Road	FR24	5	21.12423406	-156.9182066
Invasive	Molokai	Forest Reserve Road	FR25	5	21.12099546	-156.917019
Invasive	Molokai	Forest Reserve Road	FR26	5	21.11769944	-156.91843
Invasive	Molokai	Forest Reserve Road	FR27	4	21.11660561	-156.9147614
Invasive	Molokai	Forest Reserve Road	FR29	5	21.1181514	-156.9081595
Invasive	Molokai	Forest Reserve Road	FR30	5	21.11775979	-156.9224
Invasive	Molokai	Forest Reserve Road	FR31	5	21.118831	-156.9271159
Invasive	Molokai	Forest Reserve Road	FR32	5	21.1153479	-156.9213926
Invasive	Molokai	Forest Reserve Road	FR33	5	21.11443327	-156.9256379
Invasive	Molokai	Forest Reserve Road	FR34	5	NA	NA
Invasive	Molokai	Forest Reserve Road	FR35	5	21.11581184	-156.9030178
Invasive	Molokai	Molakai Kalaupapa 23	KP3	3	21.1568799	-157.0117391
Invasive	Molokai	Molakai Kalaupapa 23	KP5	5	21.17191904	-156.9989901
Invasive	Hawaii	Kahuku	KA1	5	19.06567192	-155.6782098
Invasive	Hawaii	Kahuku	KA2	2	19.06778986	-155.6782365
Invasive	Hawaii	Kahuku	KA3	5	19.07118604	-155.6791741
Invasive	Hawaii	Kahuku	KA4	5	19.07904315	-155.6807853
Invasive	Hawaii	Kahuku	KA5	5	19.08806828	-155.6844934

Table A 1 continued

Range	Region	Transect	Trap	Number	Latitude	Longitude
Invasive	Hawaii	Kahuku	KA6	6	19.09739834	-155.6873645
Invasive	Hawaii	Kahuku	KA7	4	19.10320197	-155.6920068
Invasive	Hawaii	Kahuku	KA8	5	19.10701205	-155.6979206
Invasive	Hawaii	Kahuku	KA9	3	19.11232743	-155.6958219
Invasive	Hawaii	Kahuku	KA10	4	19.1178848	-155.6960086
Invasive	Hawaii	Kahuku	KA14	1	19.14796444	-155.698604
Invasive	Hawaii	SRA	SRA1	3	19.68172973	-155.1875405
Invasive	Hawaii	SRA	SRA2	5	19.68670908	-155.1934329
Invasive	Hawaii	SRA	SRA3	2	19.6909743	-155.1992619
Invasive	Hawaii	SRA	SRA4	3	19.69316977	-155.206298
Invasive	Hawaii	SRA	SRA5	3	19.6945501	-155.2136908
Invasive	Hawaii	SRA	SRA7	3	19.69810897	-155.2271746
Invasive	Hawaii	SRA	SRA10	3	19.69451264	-155.2484585
Invasive	Kauai	Highway-552 (Kokee)	HW5	3	22.02477928	-159.6818013
Invasive	Kauai	Highway-552 (Kokee)	HW6	5	22.03280445	-159.6716323
Invasive	Kauai	Highway-552 (Kokee)	HW7	5	22.04129549	-159.6627345
Invasive	Kauai	Highway-552 (Kokee)	HW8	2	22.04890567	-159.6577262
Invasive	Kauai	Highway-552 (Kokee)	HW9	3	22.06248427	-159.664548
Invasive	Kauai	Highway-552 (Kokee)	HW10	1	22.07155298	-159.6626334
Invasive	Kauai	Highway-552 (Kokee)	HW14	5	22.10328997	-159.6761993
Invasive	Kauai	Highway-552 (Kokee)	HW16	3	22.11565428	-159.669471
Invasive	Kauai	Highway-552 (Kokee)	HW21	1	22.13824284	-159.6521541
Invasive	Kauai	Highway-552 (Kokee)	HW23	4	22.15016828	-159.6458712
Invasive	Kauai	Highway-552 (Kokee)	HW24	3	22.15079483	-159.6395576
Invasive	Kauai	Makaha Ridge	MK1	5	NA	NA
Invasive	Kauai	Makaha Ridge	MK2	1	22.11540107	-159.6757972
Invasive	Kauai	Makaha Ridge	MK3	5	22.11847689	-159.6817936
Invasive	Kauai	Makaha Ridge	MK4	5	22.12140553	-159.6890832
Invasive	Kauai	Makaha Ridge	MK5	5	22.12162572	-159.6961179
Invasive	Kauai	Makaha Ridge	MK6	5	22.12205345	-159.704097
Invasive	Kauai	Makaha Ridge	MK7	5	22.12467128	-159.7102141
Invasive	Kauai	Makaha Ridge	MK8	5	22.12784551	-159.7177472
Invasive	Kauai	Makaha Ridge	MK9	5	22.12964326	-159.721019
Invasive	Lanai	Garden of the Gods	GG1	5	20.83712253	-156.9230502
Invasive	Lanai	Monroe Trail	MR1	5	20.84831262	-156.9224792
Invasive	Lanai	Monroe Trail	MR2	4	20.84539915	-156.9197347
Invasive	Lanai	Monroe Trail	MR3	5	20.84388186	-156.9158726
Invasive	Lanai	Monroe Trail	MR4	4	20.84269054	-156.9124666
Invasive	Lanai	Monroe Trail	MR5	3	20.84272332	-156.9078532
Invasive	Lanai	Monroe Trail	MR6	3	20.84793351	-156.9068342
Invasive	Lanai	Monroe Trail	MR7	5	20.84738818	-156.9026268
Invasive	Lanai	Monroe Trail	MR9	5	20.84070772	-156.8979746

Table A 1 continued

Range	Region	Transect	Trap	Number	Latitude	Longitude
Invasive	Lanai	Monroe Trail	MR10	2	20.83652322	-156.8966233
Invasive	Lanai	Monroe Trail	MR11	4	20.83553977	-156.8962432
Invasive	Lanai	Monroe Trail	MR12	5	20.83164747	-156.8959049
Invasive	Lanai	Monroe Trail	MR13	5	20.82775617	-156.8947654
Invasive	Lanai	Monroe Trail	MR14	5	20.8257373	-156.8910627
Invasive	Lanai	Monroe Trail	MR15	5	20.82358919	-156.8874466
Invasive	Lanai	Monroe Trail	MR16	5	20.82132046	-156.8840231
Invasive	Lanai	Monroe Trail	MR17	5	20.82051907	-156.8801854
Invasive	Lanai	Monroe Trail	MR18	5	20.81773301	-156.8771285
Invasive	Lanai	Monroe Trail	MR19	5	20.81456054	-156.8750053
Invasive	Lanai	Monroe Trail	MR20	4	20.81000078	-156.8688336
Invasive	Lanai	Monroe Trail	MR21	5	20.80372743	-156.8642982
Invasive	Lanai	Monroe Trail	MR22	4	20.79681672	-156.8609913
Invasive	Lanai	Monroe Trail	MR23	5	20.7897434	-156.860762
Invasive	Lanai	Monroe Trail	MR24	3	20.78284174	-156.8641083
Invasive	Lanai	Shipwreck	SW1	1	20.85554	-156.9186545
Invasive	Maui	Hosmer Grove	HG1	1	20.76867063	-156.2376599
Invasive	Maui	Hosmer Grove	HG2	1	20.7710614	-156.2365737
Invasive	Maui	Hosmer Grove	HG3	1	20.77333658	-156.2353919
Invasive	Maui	Hosmer Grove	HG8	1	20.77504213	-156.2335384
Invasive	Maui	Haleakala	HM1	5	20.75221301	-156.2282336
Invasive	Maui	Haleakala	HM2	5	20.75285741	-156.2276546
Invasive	Maui	Haleakala	HM3	4	20.75387883	-156.2254667
Invasive	Maui	Haleakala	HM4	4	20.75548531	-156.2234869
Invasive	Maui	Haleakala	HM5	4	20.75597632	-156.2206748
Invasive	Maui	Haleakala	HM7	2	20.75626734	-156.2182254
Invasive	Maui	Haleakala	HM8	1	20.7561888	-156.2167016
Invasive	Maui	Maui Iao Valley	IV1	5	20.88290004	-156.5321929
Invasive	Maui	Maui Iao Valley	IV2	5	20.8835804	-156.5369477
Invasive	Maui	Maui Iao Valley	IV3	5	20.88109324	-156.5444043
Invasive	Maui	Maui Iao Valley	IV4	5	20.88074941	-156.5467455
Invasive	Maui	Maui Iao Valley	IV5	5	20.88042646	-156.5468379
Invasive	Maui	Maui Iao Valley	IV6	5	NA	NA
Invasive	Maui	Maui Iao Valley	IV7	5	20.87882308	-156.5495435
Invasive	Maui	Maui Iao Valley	IV8	5	NA	NA
Invasive	Maui	Maui Iao Valley	IV9	1	NA	NA
Invasive	Maui	Maui Iao Valley	IV10	4	20.87684386	-156.5533013
Invasive	Maui	Olinda Road	OL1	5	20.845872	-156.308575
Invasive	Maui	Olinda Road	OL2	5	NA	NA
Invasive	Maui	Olinda Road	OL3	5	20.832932	-156.299168
Invasive	Maui	Olinda Road	OL4	5	20.824822	-156.293546
Invasive	Maui	Olinda Road	OL5	5	20.820252	-156.292203

Table A 1 continued

Range	Region	Transect	Trap	Number	Latitude	Longitude
Invasive	Maui	Olinda Road	OL6	5	20.8133	-156.286376
Invasive	Maui	Olinda Road	OL7	4	20.808728	-156.283682
Invasive	Maui	Olinda Road	OL8	5	20.803928	-156.274757
Invasive	Maui	Olinda Road	OL9	5	20.806026	-156.279753
Invasive	Maui	Olinda Road	OL11	4	20.861104	-156.313403
Invasive	Maui	Olinda Road	OL13	4	20.876448	-156.332463
Invasive	Maui	Olinda Road	OL14	5	20.881175	-156.339662
Invasive	Maui	Olinda Road	OL15	5	20.88934	-156.34702
Invasive	Maui	Waihee Ridge Trail	WE1	5	20.95295724	-156.5313163
Invasive	Maui	Waihee Ridge Trail	WE2	5	20.95120023	-156.5341938
Invasive	Maui	Waihee Ridge Trail	WE3	4	NA	NA
Invasive	Maui	Waihee Ridge Trail	WE4	5	20.94926845	-156.5364345
Invasive	Maui	Waihee Ridge Trail	WE5	5	20.94841492	-156.5389044
Invasive	Maui	Waihee Ridge Trail	WE6	5	20.94820647	-156.542432
Invasive	Maui	Waihee Ridge Trail	WE7	6	20.9481783	-156.5446132
Invasive	Maui	Waihee Ridge Trail	WE8	5	20.94702378	-156.5463913
Invasive	Maui	Waihee Ridge Trail	WE9	5	20.94641592	-156.5494724
Invasive	Maui	Waihee Ridge Trail	WE10	5	20.94745092	-156.5511233
Invasive	Maui	Waihee Ridge Trail	WE11	4	20.94665674	-156.5517656
Invasive	Maui	Waihee Ridge Trail	WE12	5	20.95573383	-156.530095
Invasive	Maui	Waihee Ridge Trail	WE13	4	20.9580723	-156.5296333
Invasive	Maui	Waihee Ridge Trail	WE14	1	20.95807926	-156.5256083
Invasive	Maui	Waipoli Road	WI1	5	20.73762548	-156.3245379
Invasive	Maui	Waipoli Road	WI2	3	20.73403727	-156.3217367
Invasive	Maui	Waipoli Road	WI3	2	20.73157987	-156.3194131
Invasive	Maui	Waipoli Road	WI8	5	20.71323917	-156.3000113
Invasive	Maui	Waipoli Road	WI9	1	20.71052721	-156.3040014
Invasive	Maui	Waipoli Road	WI10	1	20.74121235	-156.3275549
Invasive	Maui	Waipoli Road	WI11	5	20.7456711	-156.3204174
Invasive	Maui	Waipoli Road	WI12	4	20.7488004	-156.3136812
Invasive	Maui	Waipoli Road	WI13	4	20.76429963	-156.306281
Invasive	Oahu	Manana	MA6	2	21.43420028	-157.9286596
Invasive	Oahu	Manana	MA7	4	21.4348197	-157.9267089
Invasive	Oahu	Manana	MA8	2	21.43561346	-157.9246502
Invasive	Oahu	Manana	MA9	3	21.43690118	-157.9226251
Invasive	Oahu	Manana	MA10	2	21.43769561	-157.9202892
Invasive	Oahu	Manana	MA11	5	21.43850539	-157.9187868
Invasive	Oahu	Manana	MA12	5	21.43802745	-157.9180582
Invasive	Oahu	Manana	MA13	4	21.43936303	-157.9173758
Invasive	Oahu	Manana	MA14	6	21.4394041	-157.9173216
Invasive	Oahu	Satellite Road	SR1	2	21.55225561	-158.2379216
Invasive	Oahu	Satellite Road	SR3	3	21.55478586	-158.2354298

Table A 1 continued

Range	Region	Transect	Trap	Number	Latitude	Longitude
Invasive	Oahu	Satellite Road	SR4	3	21.55807198	-158.2380137
Invasive	Oahu	Satellite Road	SR5	1	21.56058186	-158.2411111
Invasive	Oahu	Satellite Road	SR7	1	21.561527	-158.2372682
Invasive	Oahu	Satellite Road	SR9	3	21.55881956	-158.2337913
Invasive	Oahu	Satellite Road	SR10	1	21.55691352	-158.2315507
Invasive	Oahu	Satellite Road	SR12	4	21.55352664	-158.2256639
Invasive	Oahu	Satellite Road	SR13	2	21.55232065	-158.2236139
Invasive	Oahu	Satellite Road	SR14	5	21.55311911	-158.2231277
Invasive	Oahu	Satellite Road	SR16	2	21.55574768	-158.2284214
Invasive	Oahu	WV	WV1	7	21.47648337	-158.1525378
Invasive	Oahu	WV	WV2	5	21.47959691	-158.1541039
Invasive	Oahu	WV	WV4	5	21.48678993	-158.1561318
Invasive	Oahu	WV	WV5	1	21.48962494	-158.1574763
Invasive	Oahu	WV	WV6	5	21.49248015	-158.157763
Invasive	Oahu	WV	WV7	4	21.49429995	-158.1572405
Invasive	Oahu	WV	WV8	2	21.49522598	-158.1569956
Invasive	Oahu	WV	WV10	5	21.49634119	-158.1564122
Invasive	Oahu	WV	WV11	5	21.49665803	-158.156282

Table A 2 Measures of genetic diversity at microsatellite loci for V.pensylvanica from invasive regions , including number of samples (N), number of alleles (Na), effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), allelic richness (A), and number of private alleles (Np) .

Region	N	N_a	N_e	H_o	H_e	A	N_p
Molokai	29	6.786	3.186	0.637	0.662	3.013	2.000
Hawaii	18	8.500	4.321	0.709	0.723	3.531	1.000
Kauai	20	3.929	2.178	0.428	0.450	2.235	0.000
Lanai	25	6.786	3.383	0.670	0.689	3.153	1.000
Maui	57	8.857	4.282	0.696	0.716	3.394	3.000
Oahu	29	7.643	3.114	0.607	0.636	2.991	1.000

Table A 3 Assignment and Exclusion test of invasive Hawaiian populations to mainland populations in the United States. (A) Geneclass assignment scores for individuals assigned to Western (CA/OR) and Central (WY/CO/NM) combined ranges. Combined regions (West/Central) are ranked by assignment score. (B) Probability of individuals being excluded from each combined region.

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
FR7-1	West	96.415	Central	3.585	0.694	0.379
FR7-2	West	93.489	Central	6.511	0.434	0.177
FR7-3	Central	89.054	West	10.946	0.454	0.409
FR8-1	Central	55.667	West	44.333	0.857	0.827
FR8-2	West	98.799	Central	1.201	0.901	0.616
FR8-3	West	94.227	Central	5.773	0.92	0.681
FR8-4	West	98.804	Central	1.196	0.713	0.326
FR8-5	Central	78.817	West	21.183	0.91	0.882
FR9-1	West	89.241	Central	10.759	0.936	0.77
FR9-2	West	99.963	Central	0.037	0.982	0.562
FR9-3	West	89.744	Central	10.256	0.881	0.692
FR9-4	Central	62.309	West	37.691	0.275	0.213
FR9-5	West	72.994	Central	27.006	0.228	0.084
FR10-1	West	99.229	Central	0.771	0.996	0.832
FR10-2	Central	94.577	West	5.423	0.392	0.381
FR10-3	West	85.11	Central	14.89	0.715	0.579
FR10-4	West	87.204	Central	12.796	0.903	0.707
FR10-5	West	68.723	Central	31.277	0.965	0.89
FR11-1	West	97.421	Central	2.579	0.836	0.52
FR11-2	West	82.596	Central	17.404	0.752	0.527
FR11-4	West	99.358	Central	0.642	0.958	0.651
FR11-5	Central	82.122	West	17.878	0.117	0.045
FR12-1	West	97.925	Central	2.075	0.946	0.701
FR12-2	West	81.71	Central	18.29	0.994	0.952
FR12-3	Central	72.862	West	27.138	0.738	0.725
FR12-4	West	58.681	Central	41.319	0.363	0.231
FR12-5	West	87.836	Central	12.164	0.777	0.558
FR13-1	Central	57.465	West	42.535	0.679	0.551
FR13-2	West	77.401	Central	22.599	0.482	0.334
FR13-3	West	76.133	Central	23.867	0.59	0.466
FR13-4	West	98.564	Central	1.436	0.336	0.083
FR14-5	West	99.405	Central	0.595	0.962	0.664

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
FR15-1	West	80.011	Central	19.989	0.919	0.762
FR15-2	Central	88.985	West	11.015	0.907	0.932
FR15-3	Central	67.775	West	32.225	0.575	0.479
FR15-4	Central	50.871	West	49.129	0.77	0.659
FR15-5	West	99.797	Central	0.203	0.88	0.431
FR16-1	Central	98.798	West	1.202	0.374	0.547
FR16-2	West	97.931	Central	2.069	0.804	0.462
FR16-3	West	89.205	Central	10.795	0.687	0.425
FR16-4	Central	68.134	West	31.866	0.652	0.551
FR16-5	West	93.745	Central	6.255	0.748	0.459
FR17-1	West	95.89	Central	4.11	0.295	0.081
FR17-2	West	99.158	Central	0.842	0.522	0.165
FR17-3	West	92.065	Central	7.935	0.997	0.951
FR17-4	West	99.853	Central	0.147	0.655	0.194
FR17-5	West	99.872	Central	0.128	0.703	0.215
FR18-1	Central	84.834	West	15.166	0.898	0.891
FR19-1	Central	52.879	West	47.121	0.731	0.625
FR19-2	Central	66.098	West	33.902	0.94	0.941
FR19-3	Central	69.732	West	30.268	0.647	0.5
FR19-4	Central	62.076	West	37.924	0.286	0.166
FR19-5	West	99.674	Central	0.326	0.994	0.775
FR20-1	West	94.148	Central	5.852	0.815	0.536
FR20-2	West	98.476	Central	1.524	0.861	0.529
FR20-3	West	99.375	Central	0.625	0.967	0.684
FR20-4	Central	50.551	West	49.449	0.399	0.247
FR20-5	West	90.13	Central	9.87	0.867	0.636
FR21-1	Central	91.436	West	8.564	0.853	0.862
FR21-2	West	79.911	Central	20.089	0.836	0.641
FR21-3	West	94.449	Central	5.551	0.942	0.744
FR21-4	West	76.382	Central	23.618	0.376	0.183
FR21-5	West	70.609	Central	29.391	0.662	0.425
FR22-1	Central	61.057	West	38.943	0.826	0.738
FR22-2	Central	72.302	West	27.698	0.803	0.769
FR22-3	West	97.291	Central	2.709	0.892	0.604
FR22-4	West	80.974	Central	19.026	0.717	0.501
FR22-5	West	97.561	Central	2.439	0.837	0.519
FR24-1	West	73.076	Central	26.924	0.909	0.762
FR24-2	West	85.091	Central	14.909	0.735	0.507

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
FR24-3	Central	97.188	West	2.812	0.423	0.564
FR24-4	West	91.021	Central	8.979	0.431	0.214
FR24-5	Central	64.815	West	35.185	0.655	0.547
FR25-1	West	71.116	Central	28.884	0.516	0.359
FR25-2	Central	96.072	West	3.928	0.689	0.802
FR25-3	Central	87.255	West	12.745	0.876	0.91
FR25-4	West	62.57	Central	37.43	0.843	0.759
FR25-5	West	86.673	Central	13.327	0.904	0.697
FR26-1	West	65.946	Central	34.054	0.856	0.709
FR26-2	Central	79.045	West	20.955	0.632	0.555
FR26-3	West	67.304	Central	32.696	0.779	0.598
FR26-4	West	62.004	Central	37.996	0.572	0.401
FR26-5	West	99.983	Central	0.017	0.796	0.207
FR27-1	Central	56.573	West	43.427	0.935	0.889
FR27-2	West	98.231	Central	1.769	0.867	0.53
FR27-3	West	99.515	Central	0.485	0.564	0.177
FR27-4	West	93.201	Central	6.799	0.194	0.044
FR29-1	Central	73.662	West	26.338	0.484	0.387
FR29-2	West	94.87	Central	5.13	0.999	0.968
FR29-3	West	71.171	Central	28.829	0.561	0.422
FR29-4	Central	59.651	West	40.349	0.818	0.787
FR29-5	West	65.695	Central	34.305	0.552	0.416
FR30-1	Central	62.433	West	37.567	0.862	0.862
FR30-2	West	69.232	Central	30.768	0.911	0.775
FR30-3	West	96.178	Central	3.822	0.989	0.854
FR30-4	West	84.447	Central	15.553	0.546	0.351
FR30-5	West	93.575	Central	6.425	0.469	0.248
FR31-1	West	94.482	Central	5.518	0.975	0.824
FR31-2	Central	94.544	West	5.456	0.417	0.516
FR31-3	Central	90.919	West	9.081	0.077	0.015
FR31-4	West	90.657	Central	9.343	0.472	0.24
FR31-5	West	94.211	Central	5.789	0.997	0.939
FR32-1	West	97.955	Central	2.045	0.956	0.723
FR32-2	Central	93.785	West	6.215	0.611	0.644
FR32-3	West	76.835	Central	23.165	0.62	0.472
FR32-4	West	95.802	Central	4.198	0.994	0.906
FR32-5	West	99.965	Central	0.035	0.608	0.123
FR33-1	West	95.294	Central	4.706	0.812	0.571

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
FR33-2	Central	92.922	West	7.078	0.409	0.468
FR33-3	West	98.289	Central	1.711	0.862	0.533
FR33-4	West	97.024	Central	2.976	0.549	0.244
FR33-5	West	97.503	Central	2.497	0.707	0.44
FR34-1	Central	98.364	West	1.636	0.205	0.306
FR34-2	Central	54.275	West	45.725	0.577	0.5
FR34-3	West	99.989	Central	0.011	0.943	0.399
FR34-4	West	94.824	Central	5.176	0.66	0.382
FR34-5	West	68.225	Central	31.775	0.729	0.661
FR35-1	Central	79.223	West	20.777	0.834	0.849
FR35-2	Central	80.576	West	19.424	0.524	0.468
FR35-3	West	88.967	Central	11.033	0.954	0.813
FR35-4	Central	50.051	West	49.949	0.779	0.653
FR35-5	West	99.182	Central	0.818	0.751	0.346
KP3-1	West	99.942	Central	0.058	0.392	0.052
KP3-2	Central	85.181	West	14.819	0.925	0.926
KP3-3	West	75.283	Central	24.717	0.753	0.544
KP5-1	West	56.538	Central	43.462	0.523	0.44
KP5-2	Central	50.282	West	49.718	0.731	0.586
KP5-3	West	99.814	Central	0.186	0.713	0.242
KP5-4	Central	87.04	West	12.96	0.519	0.506
KP5-5	West	73.919	Central	26.081	0.42	0.265
KA1-1	West	100	Central	0	0.795	0.032
KA1-2	West	85.22	Central	14.78	0.522	0.282
KA1-3	West	99.17	Central	0.83	0.977	0.751
KA1-4	West	99.533	Central	0.467	0.554	0.168
KA1-5	West	99.855	Central	0.145	0.886	0.375
KA2-1	West	68.051	Central	31.949	0.757	0.566
KA2-2	West	97.746	Central	2.254	0.604	0.262
KA3-1	West	99.993	Central	0.007	0.793	0.177
KA3-2	West	82.069	Central	17.931	0.452	0.222
KA3-3	West	99.973	Central	0.027	0.484	0.063
KA3-4	Central	84.548	West	15.452	0.361	0.283
KA3-5	West	99.871	Central	0.129	0.975	0.607
KA4-1	West	98.895	Central	1.105	0.599	0.229
KA4-2	West	96.73	Central	3.27	0.98	0.812
KA4-3	West	99.893	Central	0.107	0.757	0.25
KA4-4	West	62.846	Central	37.154	0.816	0.655

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
KA4-5	West	80.51	Central	19.49	0.929	0.788
KA5-1	West	99.986	Central	0.014	0.435	0.04
KA5-2	West	99.828	Central	0.172	0.656	0.202
KA5-3	West	99.979	Central	0.021	0.884	0.313
KA5-4	West	99.852	Central	0.148	0.331	0.044
KA5-5	West	100	Central	0	0.578	0.037
KA6-1	West	99.945	Central	0.055	0.67	0.168
KA6-2	West	95.204	Central	4.796	0.782	0.486
KA6-3	West	96.97	Central	3.03	0.937	0.699
KA6-4	West	100	Central	0	0.996	0.415
KA6-5	West	73.098	Central	26.902	0.988	0.951
KA6-7	West	98.845	Central	1.155	0.887	0.516
KA7-1	West	99.438	Central	0.562	0.491	0.141
KA7-2	West	99.218	Central	0.782	0.807	0.372
KA7-4	West	99.977	Central	0.023	0.554	0.096
KA7-5	West	85.805	Central	14.195	0.987	0.918
KA8-1	West	99.714	Central	0.286	0.848	0.412
KA8-2	Central	50.216	West	49.784	0.583	0.44
KA8-3	West	99.816	Central	0.184	0.843	0.387
KA8-4	West	96.521	Central	3.479	0.858	0.562
KA8-5	West	99.933	Central	0.067	0.534	0.106
KA9-1	Central	62.068	West	37.932	0.435	0.304
KA9-2	West	100	Central	0	0.822	0.051
KA9-3	West	82.915	Central	17.085	0.425	0.203
KA10-1	West	100	Central	0	0.124	0.001
KA10-2	West	96.876	Central	3.124	0.565	0.249
KA10-3	Central	93.068	West	6.932	0.248	0.203
KA10-4	West	99.929	Central	0.071	0.397	0.058
KA14-1	West	99.956	Central	0.044	0.463	0.063
SRA1-1	West	89.87	Central	10.13	0.741	0.481
SRA1-2	West	87.33	Central	12.67	0.979	0.913
SRA1-3	Central	83.924	West	16.076	0.306	0.213
SRA2-1	West	91.633	Central	8.367	0.634	0.387
SRA2-2	West	99.951	Central	0.049	0.819	0.285
SRA2-3	West	100	Central	0	0.616	0.016
SRA2-4	West	98.035	Central	1.965	0.441	0.089
SRA2-5	West	99.976	Central	0.024	0.455	0.022
SRA3-1	Central	56.05	West	43.95	0.596	0.547

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
SRA3-2	West	99.966	Central	0.034	0.835	0.282
SRA4-1	West	100	Central	0	0.619	0.039
SRA4-2	Central	76.375	West	23.625	0.833	0.775
SRA4-3	West	87.516	Central	12.484	0.123	0.019
SRA5-1	West	99.834	Central	0.166	0.199	0.01
SRA5-2	West	97.296	Central	2.704	0.768	0.436
SRA5-3	West	99.433	Central	0.567	0.553	0.177
SRA7-1	West	99.058	Central	0.942	0.915	0.564
SRA7-2	West	99.979	Central	0.021	0.661	0.144
SRA7-3	West	99.998	Central	0.002	0.458	0.029
SRA10-1	West	92.786	Central	7.214	0.867	0.613
SRA10-2	Central	62.804	West	37.196	0.618	0.514
SRA10-3	West	99.998	Central	0.002	0.545	0.053
HW5-1	West	96.808	Central	3.192	0.885	0.605
HW5-2	West	70.126	Central	29.874	0.991	0.959
HW5-3	West	99.43	Central	0.57	0.991	0.835
HW6-1	West	99.402	Central	0.598	0.63	0.226
HW6-2	West	73.393	Central	26.607	0.507	0.303
HW6-3	West	90.845	Central	9.155	0.167	0.026
HW6-4	West	96.539	Central	3.461	0.701	0.382
HW6-5	West	99.975	Central	0.025	0.658	0.144
HW7-1	West	99.787	Central	0.213	0.554	0.148
HW7-2	West	98.952	Central	1.048	0.87	0.519
HW7-3	West	99.837	Central	0.163	0.792	0.304
HW7-4	West	99.397	Central	0.603	0.515	0.151
HW7-5	West	99.627	Central	0.373	0.631	0.195
HW8-2	West	99.632	Central	0.368	0.929	0.52
HW8-3	West	99.955	Central	0.045	0.657	0.156
HW9-1	West	84.193	Central	15.807	0.642	0.475
HW9-2	West	96.936	Central	3.064	0.653	0.323
HW9-3	West	99.978	Central	0.022	0.645	0.133
HW10-1	West	94.245	Central	5.755	0.787	0.45
HW14-1	Central	93.775	West	6.225	0.612	0.609
HW14-2	Central	95.255	West	4.745	0.74	0.762

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
HW14-3	Central	81.539	West	18.461	0.792	0.785
HW14-4	West	100	Central	0	0.151	0
HW14-5	West	61.066	Central	38.934	0.997	0.971
HW16-1	West	84.557	Central	15.443	0.987	0.922
HW16-2	West	99.998	Central	0.002	0.78	0.1
HW16-3	West	99.829	Central	0.171	0.488	0.107
HW21-1	West	99.178	Central	0.822	0.888	0.529
HW23-1	Central	83.612	West	16.388	0.904	0.892
HW23-2	West	95.595	Central	4.405	0.603	0.292
HW23-3	West	91.867	Central	8.133	0.624	0.346
HW23-4	West	97.023	Central	2.977	0.614	0.285
HW24-1	West	98.588	Central	1.412	0.867	0.53
HW24-2	West	99.825	Central	0.175	0.629	0.18
HW24-3	West	97.788	Central	2.212	0.614	0.266
MK1-1	West	94.872	Central	5.128	0.915	0.659
MK1-2	West	98.39	Central	1.61	0.926	0.635
MK1-3	West	99.954	Central	0.046	0.861	0.31
MK1-4	West	98.974	Central	1.026	0.633	0.249
MK1-5	West	74.465	Central	25.535	0.832	0.65
MK2-1	West	99.19	Central	0.81	0.808	0.411
MK3-1	Central	51.42	West	48.58	0.973	0.925
MK3-2	West	74.067	Central	25.933	0.89	0.788
MK3-3	West	99.924	Central	0.076	0.622	0.151
MK3-4	West	97.614	Central	2.386	0.898	0.597
MK3-5	West	95.364	Central	4.636	0.793	0.499
MK4-1	West	99.86	Central	0.14	0.946	0.553
MK4-2	West	98.827	Central	1.173	0.713	0.328
MK4-3	West	99.661	Central	0.339	0.972	0.654
MK4-4	West	99.764	Central	0.236	0.889	0.462
MK4-5	West	80.081	Central	19.919	0.865	0.674

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
MK5-1	West	99.453	Central	0.547	0.921	0.555
MK5-2	West	99.996	Central	0.004	0.814	0.17
MK5-3	West	98.652	Central	1.348	0.852	0.562
MK5-4	West	93.028	Central	6.972	0.436	0.181
MK5-5	West	99.968	Central	0.032	0.081	0.001
MK6-1	West	99.286	Central	0.714	0.454	0.127
MK6-2	West	99.98	Central	0.02	0.555	0.074
MK6-4	West	98.767	Central	1.233	0.443	0.162
MK6-5	West	74.544	Central	25.456	0.435	0.218
MK7-1	West	98.125	Central	1.875	0.637	0.282
MK7-2	West	99.989	Central	0.011	0.583	0.09
MK7-3	West	99.871	Central	0.129	0.546	0.133
MK7-4	West	99.631	Central	0.369	0.958	0.633
MK7-5	West	99.805	Central	0.195	0.315	0.048
MK8-1	West	99.905	Central	0.095	0.792	0.282
MK8-2	West	84.855	Central	15.145	0.531	0.265
MK8-3	West	98.658	Central	1.342	0.865	0.526
MK8-4	West	97.817	Central	2.183	0.726	0.379
MK8-5	West	99.986	Central	0.014	0.742	0.165
MK9-1	West	99.803	Central	0.197	0.522	0.129
MK9-2	West	99.973	Central	0.027	0.388	0.037
MK9-3	West	99.567	Central	0.433	0.843	0.426
MK9-4	West	99.2	Central	0.8	0.73	0.318
MK9-5	West	97.745	Central	2.255	0.606	0.262
GG1-1	Central	53.703	West	46.297	0.607	0.469
GG1-2	West	92.523	Central	7.477	0.128	0.017
GG1-3	West	99.998	Central	0.002	0.207	0.005
GG1-4	West	99.695	Central	0.305	0.344	0.061
GG1-5	West	76.662	Central	23.338	0.241	0.092
MR1-1	West	91.008	Central	8.992	0.834	0.551
MR1-2	West	85	Central	15	0.488	0.248
MR1-3	Central	68.146	West	31.854	0.522	0.412
MR1-4	West	99.061	Central	0.939	0.125	0.01
MR1-5	West	87.657	Central	12.343	0.546	0.248
MR2-2	West	60.329	Central	39.671	0.609	0.439
MR2-3	West	64.818	Central	35.182	0.792	0.623
MR2-4	West	73.308	Central	26.692	0.935	0.813
MR2-5	West	98.871	Central	1.129	0.804	0.499

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
MR3-1	West	99.886	Central	0.114	0.86	0.384
MR3-2	West	97.353	Central	2.647	0.742	0.397
MR3-3	West	68.095	Central	31.905	0.845	0.686
MR3-4	West	77.567	Central	22.433	0.537	0.329
MR3-5	Central	86.182	West	13.818	0.334	0.302
MR4-1	Central	87.849	West	12.151	0.547	0.518
MR4-2	West	94.944	Central	5.056	0.884	0.614
MR4-3	West	84.972	Central	15.028	0.407	0.256
MR4-5	Central	65.912	West	34.088	0.776	0.74
MR5-1	West	99.877	Central	0.123	0.329	0.039
MR5-2	West	54.026	Central	45.974	0.751	0.592
MR5-3	West	99.998	Central	0.002	0.382	0.017
MR6-2	West	85.245	Central	14.755	0.359	0.155
MR6-3	West	78.527	Central	21.473	0.862	0.726
MR6-4	Central	79.159	West	20.841	0.155	0.163
MR7-1	Central	99.895	West	0.105	0.376	0.571
MR7-2	West	99.999	Central	0.001	0.338	0.003
MR7-3	West	61.185	Central	38.815	0.164	0.04
MR7-4	Central	97.399	West	2.601	0.183	0.182
MR7-5	Central	94.81	West	5.19	0.313	0.408
MR9-1	West	99.998	Central	0.002	0.288	0.006
MR9-2	West	91.506	Central	8.494	0.793	0.53
MR9-3	West	82.074	Central	17.926	0.667	0.509
MR9-4	West	99.659	Central	0.341	0.756	0.363
MR9-5	Central	92.986	West	7.014	0.059	0.022
MR10-1	West	92.448	Central	7.552	0.887	0.646
MR10-2	West	94.337	Central	5.663	0.477	0.199
MR11-1	Central	84.546	West	15.454	0.253	0.183
MR11-2	West	92.183	Central	7.817	0.382	0.148
MR11-3	Central	70.151	West	29.849	0.17	0.123
MR11-5	West	91.206	Central	8.794	0.54	0.328
MR12-1	West	99.915	Central	0.085	0.534	0.115
MR12-2	Central	83.883	West	16.117	0.627	0.566
MR12-3	West	99.97	Central	0.03	0.293	0.016
MR12-4	Central	84.746	West	15.254	0.323	0.255
MR12-5	West	99.365	Central	0.635	0.477	0.134
MR13-1	West	82.432	Central	17.568	0.818	0.627
MR13-2	West	99.911	Central	0.089	0.818	0.31

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
MR13-3	West	99.211	Central	0.789	0.344	0.071
MR13-4	West	99.089	Central	0.911	0.99	0.814
MR13-5	West	99.99	Central	0.01	0.349	0.027
MR14-1	Central	88.785	West	11.215	0.171	0.105
MR14-2	West	99.954	Central	0.046	0.815	0.275
MR14-3	West	71.709	Central	28.291	0.353	0.168
MR14-4	West	64.626	Central	35.374	0.913	0.791
MR14-5	West	93.81	Central	6.19	0.657	0.377
MR15-1	West	96.645	Central	3.355	0.838	0.537
MR15-2	Central	95.569	West	4.431	0.35	0.422
MR15-3	West	99.877	Central	0.123	0.338	0.049
MR15-4	West	99.992	Central	0.008	0.459	0.045
MR15-5	West	99.929	Central	0.071	0.253	0.007
MR16-1	West	96.474	Central	3.526	0.267	0.072
MR16-2	West	86.826	Central	13.174	0.449	0.263
MR16-3	West	98.345	Central	1.655	0.902	0.562
MR16-4	Central	51.382	West	48.618	0.637	0.504
MR16-5	West	99.962	Central	0.038	0.522	0.106
MR17-1	West	99.97	Central	0.03	0.829	0.27
MR17-2	West	51.291	Central	48.709	0.347	0.198
MR17-3	Central	95.823	West	4.177	0.392	0.441
MR17-4	Central	76.661	West	23.339	0.853	0.822
MR17-5	Central	87.505	West	12.495	0.258	0.191
MR18-1	Central	62.754	West	37.246	0.382	0.244
MR18-2	Central	57.923	West	42.077	0.712	0.609
MR18-3	West	99.996	Central	0.004	0.841	0.202
MR18-4	West	99.969	Central	0.031	0.812	0.249
MR18-5	West	75.051	Central	24.949	0.589	0.413
MR19-1	West	99.369	Central	0.631	0.49	0.136
MR19-2	West	99.971	Central	0.029	0.135	0.002
MR19-3	West	93.129	Central	6.871	0.9	0.737
MR19-4	West	91.705	Central	8.295	0.553	0.237
MR19-5	West	99.299	Central	0.701	0.469	0.098
MR20-1	Central	98.869	West	1.131	0.259	0.295
MR20-2	West	75.95	Central	24.05	0.626	0.391
MR20-3	West	98.938	Central	1.062	0.343	0.046
MR20-4	West	95.187	Central	4.813	0.479	0.191
MR21-1	West	99.464	Central	0.536	0.792	0.372

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
MR21-2	West	94.447	Central	5.553	0.194	0.037
MR21-3	Central	65.448	West	34.552	0.31	0.23
MR21-4	Central	73.355	West	26.645	0.418	0.306
MR21-5	West	82.194	Central	17.806	0.818	0.604
MR22-1	Central	90.586	West	9.414	0.392	0.342
MR22-2	West	99.784	Central	0.216	0.248	0.029
MR22-4	West	99.733	Central	0.267	0.653	0.211
MR22-5	Central	74.152	West	25.848	0.375	0.265
MR23-1	Central	94.435	West	5.565	0.27	0.241
MR23-2	West	99.996	Central	0.004	0.527	0.056
MR23-3	West	99.181	Central	0.819	0.81	0.393
MR23-4	West	88.392	Central	11.608	0.416	0.181
MR23-5	West	99.144	Central	0.856	0.592	0.176
MR24-1	Central	86.292	West	13.708	0.244	0.174
MR24-2	West	57.896	Central	42.104	0.322	0.203
MR24-3	West	99.21	Central	0.79	0.898	0.537
SW1-1	Central	99.36	West	0.64	0.145	0.146
HG1-1	West	99.169	Central	0.831	0.801	0.403
HG2-1	West	68.428	Central	31.572	0.873	0.729
HG3-1	West	99.995	Central	0.005	0.823	0.159
HG8-1	Central	68.95	West	31.05	0.926	0.899
HM1-1	West	99.973	Central	0.027	0.451	0.053
HM1-2	West	99.981	Central	0.019	0.486	0.062
HM1-3	West	99.49	Central	0.51	0.384	0.08
HM1-4	West	56.196	Central	43.804	0.749	0.586
HM1-5	West	99.935	Central	0.065	0.386	0.052
HM2-1	West	99.998	Central	0.002	0.812	0.148
HM2-2	Central	98.144	West	1.856	0.629	0.703
HM2-3	West	99.999	Central	0.001	0.812	0.121
HM2-4	West	99.609	Central	0.391	0.707	0.227
HM2-5	West	88.877	Central	11.123	0.475	0.226
HM3-1	West	75.785	Central	24.215	0.758	0.551
HM3-2	West	82.43	Central	17.57	0.747	0.522
HM3-3	West	67.151	Central	32.849	0.778	0.595
HM3-5	West	65.594	Central	34.406	0.888	0.754
HM4-1	West	65.413	Central	34.587	0.649	0.467
HM4-2	Central	98.258	West	1.742	0.315	0.338
HM4-3	West	99.839	Central	0.161	0.961	0.586

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
HM4-5	West	97.339	Central	2.661	0.986	0.832
HM5-1	Central	93.52	West	6.48	0.655	0.656
HM5-2	West	76.374	Central	23.626	0.817	0.63
HM5-3	Central	79.322	West	20.678	0.896	0.899
HM5-5	West	95.337	Central	4.663	0.897	0.646
HM7-1	Central	52.937	West	47.063	0.614	0.478
HM7-2	West	96.312	Central	3.688	0.634	0.31
HM8-1	West	96.999	Central	3.001	0.96	0.751
IV1-1	West	86.224	Central	13.776	0.652	0.404
IV1-2	West	99.976	Central	0.024	0.336	0.028
IV1-3	West	99.687	Central	0.313	0.685	0.244
IV1-4	Central	55.25	West	44.75	0.858	0.759
IV1-5	Central	90.994	West	9.006	0.725	0.789
IV2-1	Central	64.013	West	35.987	0.778	0.665
IV2-2	Central	90.131	West	9.869	0.435	0.441
IV2-3	Central	51.579	West	48.421	0.439	0.289
IV2-4	West	96.898	Central	3.102	0.15	0.024
IV2-5	West	99.446	Central	0.554	0.893	0.519
IV3-1	Central	84.441	West	15.559	0.303	0.212
IV3-2	West	89.107	Central	10.893	0.633	0.341
IV3-3	West	96.065	Central	3.935	0.547	0.244
IV3-4	West	99.962	Central	0.038	0.445	0.038
IV3-5	West	55.547	Central	44.453	0.91	0.801
IV4-1	Central	88.166	West	11.834	0.691	0.801
IV4-2	West	81.96	Central	18.04	0.84	0.66
IV4-3	West	99.997	Central	0.003	0.46	0.024
IV4-4	West	56.091	Central	43.909	0.714	0.577
IV4-5	Central	94.573	West	5.427	0.625	0.635
IV5-1	West	67.762	Central	32.238	0.701	0.504
IV5-2	West	91.53	Central	8.47	0.571	0.296
IV5-3	Central	95.647	West	4.353	0.812	0.847
IV5-4	West	99.999	Central	0.001	0.633	0.061
IV5-5	West	99.068	Central	0.932	0.411	0.111
IV6-1	West	99.592	Central	0.408	0.809	0.381
IV6-2	West	99.941	Central	0.059	0.658	0.166
IV6-3	West	77.201	Central	22.799	0.243	0.096
IV6-4	West	99.923	Central	0.077	0.686	0.19
IV6-5	West	99.915	Central	0.085	0.812	0.294

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
IV7-1	West	99.997	Central	0.003	0.596	0.068
IV7-2	West	83.623	Central	16.377	0.924	0.762
IV7-3	West	99.292	Central	0.708	0.673	0.272
IV7-4	West	89.602	Central	10.398	0.596	0.334
IV7-5	Central	67.948	West	32.052	0.47	0.403
IV8-1	West	99.991	Central	0.009	0.546	0.068
IV8-2	West	82.25	Central	17.75	0.177	0.05
IV8-3	Central	91.146	West	8.854	0.283	0.233
IV8-4	Central	92.374	West	7.626	0.494	0.483
IV8-5	West	99.94	Central	0.06	0.123	0.002
IV9-1	West	99.999	Central	0.001	0.424	0.012
IV10-1	West	53.933	Central	46.067	0.579	0.425
IV10-2	West	64.949	Central	35.051	0.848	0.703
IV10-3	West	97.918	Central	2.082	0.534	0.203
IV10-4	Central	50.359	West	49.641	0.386	0.233
OL1-1	West	81.431	Central	18.569	0.174	0.047
OL1-2	West	78.884	Central	21.116	0.725	0.514
OL1-3	West	92.344	Central	7.656	0.476	0.186
OL1-4	Central	89.638	West	10.362	0.624	0.592
OL1-5	Central	99.942	West	0.058	0.115	0.154
OL2-1	Central	83.932	West	16.068	0.508	0.489
OL2-2	West	99.928	Central	0.072	0.779	0.256
OL2-3	Central	50.988	West	49.012	0.564	0.422
OL2-4	West	99.951	Central	0.049	0.841	0.305
OL2-5	West	97.445	Central	2.555	0.873	0.561
OL3-1	West	99.88	Central	0.12	0.796	0.292
OL3-2	West	99.482	Central	0.518	0.612	0.204
OL3-3	West	91.366	Central	8.634	0.934	0.744
OL3-4	West	87.525	Central	12.475	0.962	0.838
OL3-5	West	93.758	Central	6.242	0.815	0.537
OL4-1	Central	84.121	West	15.879	0.661	0.612
OL4-2	West	69.488	Central	30.512	0.832	0.659
OL4-3	West	93.085	Central	6.915	0.979	0.861
OL4-4	Central	87.953	West	12.047	0.713	0.675
OL4-5	West	99.673	Central	0.327	0.34	0.061
OL5-1	West	99.996	Central	0.004	0.485	0.038
OL5-2	West	99.158	Central	0.842	0.489	0.146
OL5-3	West	99.994	Central	0.006	0.387	0.027

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
OL5-4	West	90.609	Central	9.391	0.919	0.712
OL5-5	West	92.07	Central	7.93	0.269	0.084
OL6-1	West	99.997	Central	0.003	0.769	0.125
OL6-2	West	99.971	Central	0.029	0.588	0.118
OL6-3	West	99.989	Central	0.011	0.843	0.244
OL6-4	West	54.139	Central	45.861	0.595	0.498
OL6-5	West	82.121	Central	17.879	0.648	0.372
OL7-1	West	99.68	Central	0.32	0.706	0.285
OL7-2	West	95.146	Central	4.854	0.812	0.522
OL7-4	West	70.22	Central	29.78	0.616	0.419
OL7-5	West	99.7	Central	0.3	0.229	0.026
OL8-1	West	100	Central	0	0.785	0.071
OL8-2	Central	62.524	West	37.476	0.124	0.037
OL8-3	West	99.657	Central	0.343	0.392	0.068
OL8-4	West	94.148	Central	5.852	0.714	0.417
OL8-5	West	98.303	Central	1.697	0.741	0.38
OL9-1	West	92.592	Central	7.408	0.843	0.584
OL9-2	West	99.897	Central	0.103	0.801	0.292
OL9-3	West	63.423	Central	36.577	0.812	0.74
OL9-4	West	85.518	Central	14.482	0.952	0.822
OL9-5	West	93.31	Central	6.69	0.938	0.747
OL11-1	West	99.954	Central	0.046	0.552	0.112
OL11-3	Central	59.537	West	40.463	0.538	0.411
OL11-4	West	92.435	Central	7.565	0.986	0.878
OL11-5	West	99.453	Central	0.547	0.725	0.292
OL13-1	West	99.942	Central	0.058	0.509	0.092
OL13-2	West	99.988	Central	0.012	0.946	0.333
OL13-3	West	99.702	Central	0.298	0.959	0.596
OL13-4	Central	98.222	West	1.778	0.252	0.261
OL14-1	West	99.787	Central	0.213	0.275	0.035
OL14-2	West	99.975	Central	0.025	0.812	0.243
OL14-3	West	99.641	Central	0.359	0.758	0.31
OL14-4	West	99.974	Central	0.026	0.491	0.067
OL14-5	Central	92.134	West	7.866	0.199	0.162
OL15-1	West	99.822	Central	0.178	0.655	0.202
OL15-2	Central	91.432	West	8.568	0.703	0.692
OL15-3	West	99.995	Central	0.005	0.59	0.071
OL15-4	West	99.997	Central	0.003	0.786	0.146

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
OL15-5	West	88.584	Central	11.416	0.809	0.563
WE1-1	West	89.082	Central	10.918	0.757	0.509
WE1-2	West	96.663	Central	3.337	0.933	0.681
WE1-3	West	85.241	Central	14.759	0.8	0.564
WE1-4	West	99.631	Central	0.369	0.758	0.31
WE1-5	West	58.321	Central	41.679	0.594	0.431
WE2-1	Central	60.66	West	39.34	0.804	0.696
WE2-2	West	99.912	Central	0.088	0.604	0.147
WE2-3	Central	99.731	West	0.269	0.653	0.836
WE2-4	West	85.737	Central	14.263	0.745	0.514
WE2-5	West	79.968	Central	20.032	0.915	0.754
WE3-1	West	99.678	Central	0.322	0.483	0.119
WE3-2	West	99.992	Central	0.008	0.757	0.156
WE3-3	Central	71.751	West	28.249	0.429	0.313
WE3-4	West	99.913	Central	0.087	0.45	0.071
WE4-1	West	99.317	Central	0.683	0.67	0.263
WE4-2	Central	72.798	West	27.202	0.552	0.464
WE4-3	West	99.972	Central	0.028	0.804	0.239
WE4-4	West	99.953	Central	0.047	0.867	0.338
WE4-5	Central	65.926	West	34.074	0.86	0.789
WE5-1	Central	93.651	West	6.349	0.435	0.419
WE5-2	West	82.657	Central	17.343	0.382	0.174
WE5-3	West	99.339	Central	0.661	0.898	0.475
WE5-4	West	53.32	Central	46.68	0.655	0.517
WE5-5	Central	86.134	West	13.866	0.748	0.707
WE6-1	Central	80.894	West	19.106	0.253	0.174
WE6-2	West	73.414	Central	26.586	0.884	0.731
WE6-3	West	55.838	Central	44.162	0.708	0.549
WE6-4	West	99.728	Central	0.272	0.745	0.285
WE6-5	West	66.433	Central	33.567	0.712	0.531
WE7-1	Central	99.137	West	0.863	0.319	0.39
WE7-2	West	99.701	Central	0.299	0.475	0.117
WE7-3	West	95.6	Central	4.4	0.604	0.27
WE7-4	West	99.538	Central	0.462	0.225	0.021
WE7-4	West	99.779	Central	0.221	0.546	0.115
WE7-5	West	78.728	Central	21.272	0.267	0.112
WE8-1	West	99.895	Central	0.105	0.826	0.327
WE8-2	West	99.953	Central	0.047	0.67	0.165

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
WE8-3	Central	93.851	West	6.149	0.701	0.741
WE8-4	West	99.075	Central	0.925	0.843	0.472
WE8-5	West	99.796	Central	0.204	0.685	0.227
WE9-1	West	94.456	Central	5.544	0.808	0.526
WE9-2	West	99.985	Central	0.015	0.505	0.063
WE9-3	West	99.992	Central	0.008	0.653	0.116
WE9-4	Central	51.145	West	48.855	0.898	0.813
WE9-5	West	99.775	Central	0.225	0.723	0.209
WE10-1	West	99.283	Central	0.717	0.769	0.364
WE10-2	West	97.894	Central	2.106	0.448	0.107
WE10-3	West	66.126	Central	33.874	0.503	0.313
WE10-4	Central	76.031	West	23.969	0.856	0.813
WE10-5	West	69.444	Central	30.556	0.731	0.543
WE11-1	West	98.456	Central	1.544	0.934	0.643
WE11-2	West	99.478	Central	0.522	0.942	0.604
WE11-3	West	92.639	Central	7.361	0.892	0.651
WE11-5	West	99.772	Central	0.228	0.248	0.016
WE12-1	Central	88.366	West	11.634	0.988	0.987
WE12-2	West	96.991	Central	3.009	0.694	0.368
WE12-3	West	85.515	Central	14.485	0.799	0.563
WE12-4	West	98.145	Central	1.855	0.706	0.346
WE12-5	Central	96.949	West	3.051	0.173	0.164
WE13-1	West	77.855	Central	22.145	0.75	0.542
WE13-2	West	88.852	Central	11.148	0.744	0.444
WE13-3	West	93.866	Central	6.134	0.617	0.322
WE13-4	West	99.044	Central	0.956	0.982	0.75
WE14-1	West	68.093	Central	31.907	0.282	0.134
WI1-1	West	82.714	Central	17.286	0.92	0.754
WI1-2	West	90.313	Central	9.687	0.522	0.262
WI1-3	West	99.96	Central	0.04	0.73	0.191
WI1-4	West	97.601	Central	2.399	0.633	0.289
WI1-5	West	99.005	Central	0.995	0.792	0.399
WI2-1	West	99.911	Central	0.089	0.749	0.239
WI2-2	West	97.809	Central	2.191	0.892	0.584
WI2-3	West	75.744	Central	24.256	0.818	0.633
WI3-1	West	99.403	Central	0.597	0.231	0.03
WI3-4	West	98.524	Central	1.476	0.898	0.566
WI8-1	Central	70.262	West	29.738	0.769	0.671

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
WI8-2	West	83.478	Central	16.522	0.47	0.239
WI8-3	Central	59.689	West	40.311	0.333	0.202
WI8-4	West	93.793	Central	6.207	0.952	0.788
WI8-5	West	99.993	Central	0.007	0.824	0.204
WI9-1	West	97.395	Central	2.605	0.158	0.022
WI10-1	West	99.914	Central	0.086	0.358	0.041
WI11-1	West	99.979	Central	0.021	0.822	0.247
WI11-2	Central	90.324	West	9.676	0.552	0.531
WI11-3	Central	97.46	West	2.54	0.661	0.734
WI11-4	West	98.717	Central	1.283	0.496	0.123
WI11-5	Central	96.052	West	3.948	0.265	0.247
WI12-1	West	95.215	Central	4.785	0.729	0.416
WI12-2	West	99.941	Central	0.059	0.676	0.176
WI12-4	West	99.74	Central	0.26	0.453	0.088
WI12-5	West	99.767	Central	0.233	0.616	0.143
WI13-1	West	95.143	Central	4.857	0.655	0.353
WI13-2	West	96.099	Central	3.901	0.917	0.651
WI13-3	West	99.608	Central	0.392	0.511	0.142
WI13-4	West	90.836	Central	9.164	0.915	0.703
MA6-1	West	90.886	Central	9.114	0.999	0.987
MA6-2	West	99.991	Central	0.009	0.951	0.4
MA7-1	West	98.742	Central	1.258	0.766	0.354
MA7-2	West	99.866	Central	0.134	0.563	0.13
MA7-3	West	81.978	Central	18.022	0.627	0.399
MA7-4	West	99.432	Central	0.568	0.876	0.443
MA8-1	West	98.774	Central	1.226	0.897	0.54
MA8-2	West	99.113	Central	0.887	0.782	0.39
MA9-1	West	97.867	Central	2.133	0.993	0.868
MA9-2	West	99.93	Central	0.07	0.912	0.419
MA9-3	West	76.836	Central	23.164	0.954	0.857
MA10-1	West	99.91	Central	0.09	0.81	0.294
MA10-2	West	99.828	Central	0.172	0.898	0.441
MA11-1	West	99.08	Central	0.92	0.999	0.952
MA11-2	West	99.721	Central	0.279	1	0.791
MA11-3	West	96.98	Central	3.02	0.988	0.845

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
MA11-4	West	98.31	Central	1.69	0.785	0.443
MA11-5	Central	83.26	West	16.74	0.991	0.988
MA12-1	West	98.158	Central	1.842	0.686	0.395
MA12-2	West	99.742	Central	0.258	0.982	0.673
MA12-3	West	97.148	Central	2.852	1	0.983
MA12-4	West	99.977	Central	0.023	0.932	0.39
MA12-5	West	72.497	Central	27.503	0.472	0.27
MA13-1	West	95.559	Central	4.441	0.67	0.368
MA13-2	West	87.191	Central	12.809	0.815	0.579
MA13-3	Central	53.116	West	46.884	0.994	0.977
MA13-4	West	99.829	Central	0.171	1	0.911
MA14-1	West	99.551	Central	0.449	0.982	0.703
MA14-2	West	99.855	Central	0.145	0.847	0.381
MA14-3	West	99.705	Central	0.295	0.997	0.815
MA14-4	West	99.922	Central	0.078	1	0.89
MA14-5	West	99.345	Central	0.655	0.812	0.374
MA14-6	West	94.523	Central	5.477	0.556	0.265
SR1-1	West	90.186	Central	9.814	0.775	0.538
SR1-2	West	78.328	Central	21.672	0.537	0.313
SR3-1	West	100	Central	0	0.426	0.007
SR3-2	Central	97.683	West	2.317	0.429	0.476
SR3-3	West	78.13	Central	21.87	0.893	0.735
SR4-1	West	97.739	Central	2.261	0.308	0.072
SR4-2	West	93.983	Central	6.017	0.567	0.282
SR4-3	West	98.839	Central	1.161	0.995	0.847
SR5-1	West	80.575	Central	19.425	0.45	0.263
SR7-1	Central	90.601	West	9.399	0.719	0.703
SR9-1	West	78.278	Central	21.722	0.59	0.383
SR9-2	West	92.125	Central	7.875	0.959	0.832

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
SR9-3	West	99.903	Central	0.097	0.988	0.661
SR10-1	West	95.995	Central	4.005	0.846	0.528
SR12-1	West	96.324	Central	3.676	0.997	0.939
SR12-2	West	98.353	Central	1.647	1	0.938
SR12-3	West	99.906	Central	0.094	0.936	0.513
SR12-4	West	94.888	Central	5.112	0.542	0.282
SR13-1	West	99.458	Central	0.542	0.981	0.71
SR13-2	West	94.627	Central	5.373	0.792	0.507
SR14-5	West	81.933	Central	18.067	0.713	0.488
SR16-1	West	99.951	Central	0.049	0.64	0.121
SR16-2	West	98.419	Central	1.581	0.948	0.697
WV1-1	West	98.851	Central	1.149	0.933	0.625
WV1-2	West	99.452	Central	0.548	0.367	0.072
WV1-3	West	82.427	Central	17.573	0.491	0.262
WV1-4	West	97.42	Central	2.58	0.852	0.614
WV1-5	West	94.885	Central	5.115	0.787	0.435
WV1-6	West	71.03	Central	28.97	0.886	0.676
WV1-7	West	99.596	Central	0.404	0.985	0.666
WV2-1	West	99.343	Central	0.657	0.986	0.762
WV2-2	Central	63.951	West	36.049	0.731	0.621
WV2-3	West	97.523	Central	2.477	0.981	0.822
WV2-4	West	94.672	Central	5.328	0.952	0.762
WV2-5	West	94.458	Central	5.542	0.715	0.415
WV4-1	West	97.245	Central	2.755	0.954	0.735
WV4-2	West	99.643	Central	0.357	0.441	0.125
WV4-3	West	99.462	Central	0.538	0.992	0.795
WV4-4	West	99.915	Central	0.085	0.54	0.101
WV4-5	West	86.031	Central	13.969	0.893	0.673
WV5-1	West	99.918	Central	0.082	0.948	0.471
WV6-1	West	96.399	Central	3.601	0.996	0.907
WV6-2	West	87.827	Central	12.173	0.888	0.672
WV6-3	West	94.778	Central	5.222	0.91	0.688
WV6-4	West	96.667	Central	3.333	0.999	0.939
WV6-5	West	99.318	Central	0.682	0.906	0.519
WV7-1	West	92.058	Central	7.942	0.997	0.928
WV7-2	West	96.537	Central	3.463	0.772	0.441
WV7-3	West	99.259	Central	0.741	0.937	0.649
WV7-5	West	99.616	Central	0.384	0.988	0.761

Table A 3 continued

A.					B.	
Assignment of Individuals (Threshold: 0.05)					**Exclusion of Individuals (Threshold: ≥ 0.01 can't be excluded from source population)**	
Sample	**rank 1**	**Assignment score [%]**	**rank 2**	**Assignment score [%]**	**West probability**	**Central probability**
WV8-1	West	99.596	Central	0.404	0.997	0.815
WV8-2	West	99.963	Central	0.037	0.676	0.159
WV10-1	Central	55.932	West	44.068	0.983	0.962
WV10-2	West	95.939	Central	4.061	0.858	0.569
WV10-3	West	98.784	Central	1.216	0.986	0.796
WV10-4	West	99.167	Central	0.833	0.756	0.348
WV10-5	West	94.575	Central	5.425	0.91	0.654
WV11-1	West	99.313	Central	0.687	0.573	0.193
WV11-2	West	88.645	Central	11.355	0.819	0.579
WV11-3	West	99.198	Central	0.802	0.969	0.668
WV11-4	West	99.674	Central	0.326	0.882	0.446
WV11-5	West	99.93	Central	0.07	0.537	0.141

APPENDIX B. SUPPLEMENT MATERIAL FOR CHAPTER 3

Table B 1 New microsatellite loci developed for *H. glaber*.

Locus	Sequence (5' to 3')	Repeat	Size (bp)	T_m	PCR Method
Hgla_243.1	F: CTACTGAGCTGCTTCGAGCC R: TGCAGAAGTCATCCTTGGCA	$(AC)_9$	249	69	Traditional
Hgla_330	F: AGGGTTTTCCCAGTCACGACGTTACCTGTCTGTGTGCATGTGT R: CAAGCACACACCTGGAGCTA	$(TG)_8$	234	68	m13
Hgla_857	F: AGGGTTTTCCCAGTCACGACGTTTGTCTTGGTGCCCACTTACC R: TCACATGATGGCAACTGGCT	$(GA)_6$	252	68	Traditional
Hgla_2663	F: AGGGTTTTCCCAGTCACGACGTTCCCACTCCATCTCTCAAGGC R: TGCCTGTAATCCCAACAGCT	$(TTGT)_7$	263	68	m13
Hgla_2681	F: CCCATGATCACAGCGAGACA R: AGTTTGCCCTCCAGTTTCCT	$(AATA)_5$	254	68	Traditional
Hgla_2793	F: AGGGTTTTCCCAGTCACGACGTTACAGAGAGAGGGAGAGAAAGAGA R: TGTGTGCTGAAGATGACATCCA	$(AG)_{14}$	220	67	m13
Hgla_3223	F: AGGGTTTTCCCAGTCACGACGTTCCCCACCTACCCACCTATGA R: TGGATTCTGGTGTGGGTTCA	$(AT)_8$	196	66	m13
Hgla_3322	F: AGGGTTTTCCCAGTCACGACGTTTGTTCTAACACAGTTAAGTTGACTTCA R: ACACAGATTCACAAAACTGTTAGCA	$(TA)_{14}$	280	66 67	m13
Hgla_3591	F: AGGGTTTTCCCAGTCACGACGTTTCACTGACTGCAACCATAGGT R: TGCTAATGTTTAACAACTAGCTTTCCA	$(AT)_{16}$	254	67	m13

Table B 1 continued

Hgla_4233.1	F: AGGGTTTTCCCAGTCACGACGTTAGCCGCCAACTGTGAACTAA R: AGTAAGTACCATTTGACAAAAAGCT	$(AC)_{11}$	236	66	m13
Hgla_4598	F: AGGGTTTTCCCAGTCACGACGTTATGACACAATGCAGGGGAGG R: AGGCAGTGGCACAAGATGAA	$(GT)_6$	231	67	m13
Hgla_4642	F: AGGGTTTTCCCAGTCACGACGTTGCGGGGCATTTGTTTCCTTT R: AACTCAGGACCTCGTGCTTG	$(TG)_7$	231	71	m13
Hgla_6197	F: AGGGTTTTCCCAGTCACGACGTTGCGGACCCTAAATCTGGCTT R: ACACCATGCTCACACACACA	$(TG)_{13}$	276	68	m13
Hgla_6226	F: AGGGTTTTCCCAGTCACGACGTTAAATGCAGTGTTTGGCAGGG R: GCACCCACTGCTTGTCTGTA	$(AC)_8$	264	71	m13
Hgla_6655	F: AGGGTTTTCCCAGTCACGACGTTTCTGTGCACGTACCAACTCC R: TGTGGACCCTGATGCATGAC	$(GT)_6$	240	71	m13
Hgla_6757.2	F: AGGGTTTTCCCAGTCACGACGTTAATCTCTCTCCCCCAGCTGT R: TATTGGATGACACCCGGCAG	$(AC)_{15}$	168	67	m13
Hgla_7076	F: GGCTTGGCCTGAACTGTGTA R: TCAGTGAGCATCTTGTACAAGTGA	$(GT)_7$	157	66	Traditional
Hgla_7146	F: AGGGTTTTCCCAGTCACGACGTTGGCGGGAGTAATGGACACAG R: CAACATGCCTGGCTGGAAAC	$(CT)_7$	215	69	m13
Hgla_7221.2	F: TCAACTGTCTGGGATCCCCT R: CTGTGGCCCTTGGAACAGTA	$(CA)_{13}$	209	66	Traditional
Hgla_7269	F: CCCAGAGGACACACTGAAAGA R: CCACCTGTCTCAGCCTCCTA	$(TA)_6$	243	68	Traditional
Hgla_7285	F: AGGGTTTTCCCAGTCACGACGTTGCTTTGCTCTTGTTGCCCAA R: GCTCAGTGGTTCTGCTGAGT	$(TG)_7$	205	67	m13

Table B 1 continued

Hgla_7633	F: AGGGTTTTCCCAGTCACGACGTTAAGTGAGAACATACACCCATGT R: GACCGGGAGAGCTAGAATGC	$(TC)_9$	195	71	m13
Hgla_7797.1	F: AAGTGAGAACATACACCCATGT R: GACCGGGAGAGCTAGAATGC	$(GT)_6$	144	68	Traditional
Hgla_7797.2	F: GCATTCTAGCTCTCCCGGTC R: TTCTGGAGGGATAGGTGGCA	$(TC)_6$	277	66	Traditional
Hgla_7804	F: CGTGTCCTCTTGGTGTGACA R: ACAGTCTGCCTTCACGATCG	$(AC)_{14}$	110	66	Traditional
Hgla_7996	F: TCACAAGCACAAGGTCCCAG R: CTCCTCCCTTGATCCCTCCA	$(AATA)_5$	200	66	Traditional
Hgla_8448.2	F: GGGCTTCTTCACCCAACAGT R: GCCAGCCTGAGATCCTGTTT	$(ATTT)_5$	198	66	Traditional
Hgla_9217	F: ACTGTGACGTGATAAAGTGGCT R: CAGTAGCAGAGCCTGAGCAT	$(TA)_8$	181	68	Traditional
Hgla_9338	F: AGGGTTTTCCCAGTCACGACGTTTCTGTGGTCTTTCTCACACAC R: TGACAAAGTTGGACTATGCACA	$(AC)_7$	217	67	m13
Hgla_9415	F: AGGGTTTTCCCAGTCACGACGTTTGCCGAGAAGGTGCAGAAAT R: GCCTGGGCAAACTAGTGAGA	$(TC)_8$	258	69	m13
Hgla_10012	F: AGGGTTTTCCCAGTCACGACGTTGATTTCTAGTGTGCACGCGC R: GCAAGTTCAAGCCCACCATG	$(TG)_7$	146	71	m13
Hgla_10193	F: AGTGATAAGGGGCTGGGGAT R: GTTCAAGCCCAAGCCACATG	$(AC)_7$	181	68	Traditional

T_m annealing temperature, PCR method is the PCR protocol used for the specified set of primers; see main text for details.

Table B 2 Collection dates and sexes for all analyzed *H. glaber* individuals.

Zoo	Collection Date	Sex
Zoo Atlanta, Atlanta, GA	May 29, 2013	M
	May 29, 2013	M
	May 29, 2013	F
	May 29, 2013	F
	May 29, 2013	F
	May 29, 2013	M
	January 25, 2014	F
	January 25, 2014	F
	January 25, 2014	F
	January 25, 2014	F
	January 25, 2014	M
	January 25, 2014	F
	January 25, 2014	M
	January 25, 2014	M
	January 25, 2014	M
	January 25, 2014	F
	January 25, 2014	F
	January 25, 2014	M
	January 25, 2014	M
	January 25, 2014	M
	January 25, 2014	M
	January 25, 2014	M
	January 25, 2014	F
	January 25, 2014	M
	January 25, 2014	M
	January 25, 2014	F
	April 17, 2014	M
	April 17, 2014	F
	April 17, 2014	M
	April 17, 2014	M
	April 17, 2014	M
	April 17, 2014	F
	April 17, 2014	M
	April 17, 2014	F
	April 17, 2014	M
	April 17, 2014	M
	April 17, 2014	F
	April 17, 2014	M
	April 17, 2014	F
	April 17, 2014	M
	August 26, 2014	M
	August 26, 2014	M
	August 26, 2014	F
	August 26, 2014	F
	August 26, 2014	F
	August 26, 2014	F

Table B 2 continued

	August 26, 2014	F
	August 26, 2014	M
	July 11, 2014	M
	July 11, 2014	M
	July 11, 2014	M
	July 11, 2014	M
	July 11, 2014	F
	July 11, 2014	M
	July 11, 2014	M
	July 11, 2014	M
	July 11, 2014	F
	July 11, 2014	F
	July 11, 2014	F
	July 11, 2014	M
Smithsonian National Zoological Park, Washington, D.C.	May 24, 2015	M
	May 24, 2015	F
	May 24, 2015	M
	May 24, 2015	M
	May 24, 2015	F
	May 24, 2015	M
	Unknown	M
	Unknown	M
	Unknown	M
	Unknown	F
	Unknown	F
	November 20, 2015	Unknown
	November 4, 2015	Unknown
	November 4, 2015	Unknown
	November 4, 2015	Unknown
	November 4, 2015	Unknown
	November 4, 2015	Unknown
	November 4, 2015	Unknown
	November 4, 2015	Unknown
	November 4, 2015	Unknown
	November 4, 2015	Unknown
San Diego Zoo, San Diego, CA	May 2, 2012	F
	April 7, 2013	F
	December 27, 2013	F
	May 3, 2012	F
	December 27, 2013	F
	November 7, 2010	F
	December 20, 2012	F
	November 7, 2010	M
	April 7, 2013	M
	April 7, 2013	M
	April 7, 2013	M

APPENDIX C. SUPPLEMENT MATERIAL FOR CHAPTER 4

Table C 1 Sequencing and Alignment Statistics for RNAseq

Species	Dataset	Run	Caste	Body Part/Life stage	Number of sequenced Reads	Average Spot Length	Overall alignment rate	Library Layout	N0
A. mellifera	Vleurinck et al	SRR2954344	Queen	p3	2.18E+08	101	72.32%	Single	7258879
A. mellifera	Vleurinck et al	SRR2954345	Queen	p3	2.21E+08	101	73.96%	Single	12175171
A. mellifera	Vleurinck et al	SRR2954346	Queen	p3	2.15E+08	101	74.40%	Single	8076411
A. mellifera	Vleurinck et al	SRR789759	Worker	P4	1.76E+08	101	74.15%	Single	43902465
A. mellifera	Vleurinck et al	SRR789760	Worker	P4	1.68E+08	101	75.68%	Single	91029544
A. mellifera	Vleurinck et al	SRR789761	Worker	P4	1.79E+08	101	74.83%	Single	35686652
A. mellifera	Vleurinck et al	SRR789762	drone	P4	1.89E+08	101	73.85%	Single	33648051
A. mellifera	Vleurinck et al	SRR789763	drone	P4	1.63E+08	101	74.03%	Single	40667326
A. mellifera	Vleurinck et al	SRR789764	drone	P4	1.54E+08	101	74.30%	Single	37833232
A. mellifera	Jasper et al	SRR1239302	worker	Antennae (Forager)	32055165	200	33.43%	Paired	19441512
A. mellifera	Jasper et al	SRR1239303	worker	Antennae (Forager)	36680995	200	59.91%	Paired	13172143
A. mellifera	Jasper et al	SRR1239304	worker	Antennae (Forager)	34093536	200	60.15%	Paired	12119061
A. mellifera	Jasper et al	SRR1239305	worker	Antennae (Nurse)	34029091	200	60.29%	Paired	11393061

Table C1 continued

Species	Dataset	Run	Caste	Body Part/Life stage	Number of sequenced Reads	Average Spot Length	Overall alignment rate	Library Layout	N0
A. mellifera	Jasper et al	SRR12 39306	worker	Antennae (Nurse)	33920814	200	60.29%	Paired	1208 1459
A. mellifera	Jasper et al	SRR12 39307	worker	Antennae (Nurse)	39385296	200	44.88%	Paired	1964 9878
A. mellifera	Jasper et al	SRR12 39308	worker	Midgut (Forager)	21259998	200	70.29%	Paired	5679 205
A. mellifera	Jasper et al	SRR12 39309	worker	Midgut (Forager)	22664964	200	61.49%	Paired	7997 938
A. mellifera	Jasper et al	SRR12 39310	worker	Midgut (Forager)	27272788	200	65.66%	Paired	8475 868
A. mellifera	Jasper et al	SRR12 39311	worker	Midgut - Nurse	25767228	200	65.66%	Paired	1371 3604
A. mellifera	Jasper et al	SRR12 39312	worker	Midgut - Nurse	21782475	200	59.93%	Paired	7899 511
A. mellifera	Jasper et al	SRR12 39313	worker	Midgut - Nurse	37729366	200	67.59%	Paired	1081 4898
A. mellifera	Jasper et al	SRR12 54946	worker	Hypopharyngeal gland (Forager)	42348249	199	75.70%	Paired	9762 680
A. mellifera	Jasper et al	SRR12 54947	worker	Hypopharyngeal gland (Forager)	35802042	199	76.08%	Paired	8089 838
A. mellifera	Jasper et al	SRR12 54948	worker	Hypopharyngeal gland (Forager)	50385644	199	75.52%	Paired	1169 4773
A. mellifera	Jasper et al	SRR12 54950	worker	Hypopharyngeal gland (Nurse)	34076128	199	69.84%	Paired	9824 813
A. mellifera	Jasper et al	SRR12 54951	worker	Hypopharyngeal gland (Nurse)	49947491	199	73.30%	Paired	1257 8391
A. mellifera	Jasper et al	SRR12 54952	worker	Hypopharyngeal gland (Nurse)	43507227	199	80.07%	Paired	8218 457
A. mellifera	Jasper et al	SRR12 54954	worker	Malpighian tubule (Forager)	51093091	199	63.45%	Paired	1773 6717
A. mellifera	Jasper et al	SRR12 54956	worker	Malpighian tubule (Forager)	47900044	199	63.77%	Paired	1657 5208
A. mellifera	Jasper et al	SRR12 54957	worker	Malpighian tubule (Forager)	39127169	199	59.29%	Paired	1513 1759

Table C1 continued

Species	Dataset	Run	Caste	Body Part/Life stage	Number of sequenced Reads	Average Spot Length	Overall alignment rate	Library Layout	N0
A. mellifera	Jasper et al	SRR12 54958	worker	Malpighian tubule (Nurse)	50224721	199	55.79%	Paired	2100 3561
A. mellifera	Jasper et al	SRR12 54959	worker	Malpighian tubule (Nurse)	44410724	199	62.55%	Paired	1571 0118
A. mellifera	Jasper et al	SRR12 54960	worker	Malpighian tubule (Nurse)	42736881	199	56.22%	Paired	1781 0020
A. mellifera	Jasper et al	SRR12 55009	worker	Mandibular gland (Forager)	28171454	200	63.70%	Paired	8704 563
A. mellifera	Jasper et al	SRR12 55010	worker	Mandibular gland (Forager)	40263782	200	61.54%	Paired	1296 4933
A. mellifera	Jasper et al	SRR12 55011	worker	Mandibular gland (Forager)	56687415	200	62.13%	Paired	1822 8435
A. mellifera	Jasper et al	SRR12 55012	worker	Mandibular gland (Nurse)	29524694	200	56.55%	Paired	1071 6661
A. mellifera	Jasper et al	SRR12 55013	worker	Mandibular gland (Nurse)	31987604	200	62.82%	Paired	9894 155
A. mellifera	Jasper et al	SRR12 55014	worker	Mandibular gland (Nurse)	58195535	200	57.09%	Paired	2155 1162
A. mellifera	Jasper et al	SRR12 55064	worker	Muscle (Forager)	52425201	200	66.21%	Paired	1579 4061
A. mellifera	Jasper et al	SRR12 55065	worker	Muscle (Forager)	34498791	200	67.67%	Paired	9960 287
A. mellifera	Jasper et al	SRR12 55066	worker	Muscle (Forager)	30478929	200	68.58%	Paired	8577 595
A. mellifera	Jasper et al	SRR12 55068	worker	Muscle (Nurse)	36391013	200	28.54%	Paired	2343 8418
A. mellifera	Jasper et al	SRR12 55149	worker	Muscle (Nurse)	34059175	200	39.73%	Paired	1825 6747
A. mellifera	Jasper et al	SRR12 55150	worker	Muscle (Nurse)	29070870	200	29.20%	Paired	1851 5354
A. mellifera	Jasper et al	SRR12 55151	worker	Nasonov gland (Forager)	29565877	200	67.71%	Paired	8563 969
A. mellifera	Jasper et al	SRR12 55152	worker	Nasonov gland (Forager)	34250042	200	65.62%	Paired	1054 5767

Table C1 continued

Species	Dataset	Run	Caste	Body Part/Life stage	Number of sequenced Reads	Average Spot Length	Overall alignment rate	Library Layout	N0
A. mellifera	Jasper et al	SRR1255153	worker	Nasonov gland (Forager)	35945103	200	64.17%	Paired	11532856
A. mellifera	Jasper et al	SRR1255154	worker	Nasonov gland (Nurse)	29372860	200	41.38%	Paired	15625493
A. mellifera	Jasper et al	SRR1255260	worker	Nasonov gland (Nurse)	43022527	200	65.82%	Paired	13168687
A. mellifera	Jasper et al	SRR1255326	worker	Nasonov gland (Nurse)	35499113	200	55.72%	Paired	14184572
A. mellifera	Jasper et al	SRR1255456	worker	Sting Gland (Nurse)	26549254	200	28.70%	Paired	17123746
A. mellifera	Jasper et al	SRR1255541	worker	Second Thoracic Ganglia (Forager)	34339445	200	58.60%	Paired	12395093
A. mellifera	Jasper et al	SRR1255542	worker	Second Thoracic Ganglia (Forager)	36944787	200	59.40%	Paired	13097313
A. mellifera	Jasper et al	SRR1255543	worker	Second Thoracic Ganglia (Forager)	34784343	200	59.12%	Paired	12366154
A. mellifera	Jasper et al	SRR1255544	worker	Second Thoracic Ganglia (Nurse)	32555322	200	61.32%	Paired	11164517
A. mellifera	Jasper et al	SRR1255545	worker	Second Thoracic Ganglia (Nurse)	37378431	200	58.42%	Paired	13465679
A. mellifera	Jasper et al	SRR1255546	worker	Second Thoracic Ganglia (Nurse)	31677666	200	52.96%	Paired	12954445
A. mellifera	Jasper et al	SRR1269199	worker	Sting Gland (Forager)	23657316	200	64.63%	Paired	7532491
A. mellifera	Cameron et al	SRR1028781	Queen	Larvae	7344483	50	77.14%	Single	16558853
A. mellifera	Cameron et al	SRR1028782	Queen	Larvae	6968503	50	77.01%	Single	15790542
A. mellifera	Cameron et al	SRR1028783	Worker	Larvae	7326580	50	80.25%	Single	14240234
A. mellifera	Cameron et al	SRR1028784	Worker	Larvae	7371304	50	76.83%	Single	16830631
A. mellifera	Ashby et al	SRR1571716	drone	Larvae L5	12353487	275	39.82%	Paired	58960692

Table C1 continued

Species	Dataset	Run	Caste	Body Part/Life stage	Number of sequenced Reads	Average Spot Length	Overall alignment rate	Library Layout	N0
A. mellifera	Ashby et al	SRR1571717	drone	Larvae L6	7797451	275	39.65%	Paired	3613128
A. mellifera	Ashby et al	SRR1571719	drone	Larvae L7	7440637	275	41.54%	Paired	3554727
A. mellifera	Ashby et al	SRR1571720	drone	Larvae L8	30081701	275	41.81%	Paired	14134427
A. mellifera	Ashby et al	SRR1571721	drone	Larvae L9	14657321	275	41.48%	Paired	6767484
A. mellifera	Ashby et al	SRR1571722	Queen	Larvae L10	10164930	275	32.59%	Paired	5095636
A. mellifera	Ashby et al	SRR1571723	Queen	Larvae L11	4512205	275	33.48%	Paired	2409908
A. mellifera	Ashby et al	SRR1571724	Queen	Larvae L12	5616574	275	30.76%	Paired	2475022
A. mellifera	Ashby et al	SRR1571725	Queen	Larvae L13	6458764	275	27.73%	Paired	2684413
A. mellifera	Ashby et al	SRR1571726	Queen	Larvae L14	6185263	275	29.30%	Paired	2796401
A. mellifera	Ashby et al	SRR1571727	Worker	Larvae L15	5270871	275	37.70%	Paired	2007547
A. mellifera	Ashby et al	SRR1571728	Worker	Larvae L16	6794183	275	36.51%	Paired	2626314
A. mellifera	Ashby et al	SRR1571729	Worker	Larvae L17	9219846	277	43.14%	Paired	2885854
A. mellifera	Ashby et al	SRR1571730	Worker	Larvae L18	7148814	277	43.01%	Paired	2115339
A. mellifera	Ashby et al	SRR1571731	Worker	Larvae L19	9195547	277	42.62%	Paired	2726208
Bombus terrestris audax	Harrison et al	ERR883767	Mother Queen	Whole body	19228973	50	88.07%	Single	2266326
Bombus terrestris audax	Harrison et al	ERR883768	Mother Queen	Whole body	17832666	50	88.07%	Single	2113267

Table C1 continued

Species	Dataset	Run	Caste	Body Part/Life stage	Number of sequenced Reads	Average Spot Length	Overall alignment rate	Library Layout	N0
Bombus terrestris audax	Harrison et al	ERR883769	Mother Queen	Whole body	20337107	50	88.80%	Single	2261070
Bombus terrestris audax	Harrison et al	ERR883770	Male Adult	Whole body	18829975	50	83.81%	Single	3024242
Bombus terrestris audax	Harrison et al	ERR883771	Male Adult	Whole body	14589067	50	83.05%	Single	2453417
Bombus terrestris audax	Harrison et al	ERR883772	Male Adult	Whole body	14720280	50	83.28%	Single	2445610
Bombus terrestris audax	Harrison et al	ERR883773	Male Larvae	Whole body	16266154	50	88.66%	Single	1835166
Bombus terrestris audax	Harrison et al	ERR883774	Male Larvae	Whole body	14603360	50	89.67%	Single	1501644
Bombus terrestris audax	Harrison et al	ERR883775	Male Larvae	Whole body	16310880	50	88.78%	Single	1821639
Bombus terrestris audax	Harrison et al	ERR883776	Male Pupae	Whole body	19235640	50	81.97%	Single	3445922
Bombus terrestris audax	Harrison et al	ERR883777	Male Pupae	Whole body	16800890	50	80.61%	Single	3232495
Bombus terrestris audax	Harrison et al	ERR883778	Male Pupae	Whole body	15488784	50	86.06%	Single	2146071
Bombus terrestris audax	Harrison et al	ERR883779	Queen Larvae	Whole body	14338587	50	90.33%	Single	1378574

Table C1 continued

Species	Dataset	Run	Caste	Body Part/Life stage	Number of sequenced Reads	Average Spot Length	Overall alignment rate	Library Layout	N0
Bombus terrestris audax	Harrison et al	ERR883780	Queen Pupae	Whole body	17780019	50	85.52%	Single	2557179
Bombus terrestris audax	Harrison et al	ERR883781	Virgin Queen	Whole body	18791069	50	83.35%	Single	3111661
Bombus terrestris audax	Harrison et al	ERR883782	Worker Reproductive Adult	Whole body	13855780	50	87.71%	Single	1691268
Bombus terrestris audax	Harrison et al	ERR883783	Worker Reproductive Adult	Whole body	20696845	50	86.99%	Single	2679205
Bombus terrestris audax	Harrison et al	ERR883784	Worker Reproductive Adult	Whole body	17339549	50	88.87%	Single	1919379
Bombus terrestris audax	Harrison et al	ERR883785	Worker Adult	Whole body	16057607	50	87.53%	Single	1991998
Bombus terrestris audax	Harrison et al	ERR883786	Worker Adult	Whole body	17407931	50	84.78%	Single	2635018
Bombus terrestris audax	Harrison et al	ERR883787	Worker Adult	Whole body	17962612	50	73.70%	Single	4696388
Bombus terrestris audax	Harrison et al	ERR883788	Worker Larvae	Whole body	16388426	50	77.85%	Single	3590083
Bombus terrestris audax	Harrison et al	ERR883789	Worker Larvae	Whole body	17050136	50	89.43%	Single	1789305
Bombus terrestris audax	Harrison et al	ERR883790	Worker Larvae	Whole body	17154058	50	88.94%	Single	1881286

Table C1 continued

Species	Dataset	Run	Caste	Body Part/Life stage	Number of sequenced Reads	Average Spot Length	Overall alignment rate	Library Layout	N0
Bombus terrestris audax	Harrison et al	ERR883791	Worker Pupae	Whole body	16956446	50	86.12%	Single	2338384
Bombus terrestris audax	Harrison et al	ERR883792	Worker Pupae	Whole body	23722321	50	88.87%	Single	2621291
Bombus terrestris audax	Harrison et al	ERR883793	Worker Pupae	Whole body	19589338	50	85.07%	Single	2906894

N0, RSEM: Number of reads failed to align

REFERENCES

Akre, R. D., A. Greene, J. F. MacDonald, P. J. Landolt, and H. G. Davis. 1981. The yellowjackets of America north of Mexico. Pp. 1-102. U S Department of Agriculture Agriculture Handbook.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Ascunce, M. S., C. C. Yang, J. Oakey, L. Calcaterra, W. J. Wu, C. J. Shih, J. Goudet, K. G. Ross, and D. Shoemaker. 2011. Global invasion history of the fire ant *Solenopsis invicta*. Science 331:1066-1068.

Ashby, R., S. Forêt, I. Searle, and R. Maleszka. 2016. MicroRNAs in honey bee caste determination. Pp. 18794. Scientific Reports. The Author(s).

Assis, R. and D. Bachtrog. 2013. Neofunctionalization of young duplicate genes in Drosophila. Proc. Natl. Acad. Sci. U. S. A. 110:17409-17414.

Assis, R. and D. Bachtrog. 2015. Rapid divergence and diversification of mammalian duplicate gene functions. BMC Evol. Biol. 15:138.

Atallah, J., D. C. Plachetzki, W. C. Jasper, and B. R. Johnson. 2013. The utility of shallow RNA-Seq for documenting differential gene expression in genes with high and low levels of expression. PLoS One.

Auger-Rozenberg, M. A., T. Boivin, E. Magnoux, C. Courtin, A. Roques, and C. Kerdelhue. 2012. Inferences on population history of a seed chalcid wasp: invasion success despite a severe founder effect from an unexpected source population. 21:6086-6103.

Batra, S. 1966. Nests and social behavior of halictine bees of India (Hymenoptera: Halictidae). 28:375-393.

Beaumont, M. A. 2010. Approximate Bayesian Computation in Evolution and Ecology. 41:379-406.

Begall, S. and H. Burda. 1998. Reproductive characteristics and growth in the eusocial Zambian Common mole-rat (Cryptomys sp., Bathyergidae). Z. Saugetierkd.-Int. J. Mamm. Biol. 63:297-306.

Beggs, J. R., E. G. Brockerhoff, J. C. Corley, M. Kenis, M. Masciocchi, F. Muller, Q. Rome, and C. Villemant. 2011. Ecological effects and management of invasive alien Vespidae. Biocontrol 56:505-526.

Benjamini, Y. and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. 57:289-300.

Bennett, N. C. 1990. Behaviour and social organization in a colony of the Damaraland mole-rat Cryptomys damarensis. 220:225-247.

Bennett, N. C. and C. G. Faulkes. 2000. African mole-rats: ecology and eusociality. Cambridge University Press.

Berens, A. J., J. H. Hunt, and A. L. Toth. 2015. Comparative Transcriptomics of Convergent Evolution: Different Genes but Conserved Pathways Underlie Caste Phenotypes across Lineages of Eusocial Insects. Mol. Biol. Evol. 32:690-703.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Bonasio, R., Q. Li, J. Lian, N. S. Mutti, L. Jin, H. Zhao, P. Zhang, P. Wen, H. Xiang, Y. Ding, Z. Jin, S. S. Shen, Z. Wang, W. Wang, J. Wang, S. L. Berger, J. Liebig, G. Zhang, and D. Reinberg. 2012. Genome-wide and Caste-Specific DNA Methylomes of the Ants *Camponotus floridanus* and *Harpegnathos saltator*. 22:1755-1764.

Bonasio, R., G. J. Zhang, C. Y. Ye, N. S. Mutti, X. D. Fang, N. Qin, G. Donahue, P. C. Yang, Q. Y. Li, C. Li, P. Zhang, Z. Y. Huang, S. L. Berger, D. Reinberg, J. Wang, and J. Liebig. 2010. Genomic Comparison of the Ants *Camponotus floridanus* and *Harpegnathos saltator*. Science 329:1068-1071.

Bonduriansky, R. and S. F. Chenoweth. 2009. Intralocus sexual conflict. Trends Ecol. Evol. 24:280-288.

Boomsma, J. J. and A. Grafen. 1991. Colony-level sex-ratio selection in the eusocial Hymenoptera. J. Evol. Biol. 4:383-407.

Bourke, A. F. G. 2015. Sex investment ratios in eusocial Hymenoptera support inclusive fitness theory. 28:2106-2111.

Bourke AFG, F. N. 1995. Social Evolution in Ants. Princeton University Press, Princeton, New Jersey.

Braude, S. 2000. Dispersal and new colony formation in wild naked mole-rats: evidence against inbreeding as the system of mating. Behav. Ecol. 11:7-12.

Brockerhoff, E. G., B. I. P. Barratt, J. R. Beggs, L. L. Fagan, M. K. Kay, C. B. Phillips, and C. J. Vink. 2010. Impacts of exotic invertebrates on New Zealand's indigenous species and ecosystems. N. Z. J. Ecol. 34:158-174.

Burland, T. M., N. C. Bennett, J. U. M. Jarvis, and C. G. Faulkes. 2002. Eusociality in African Mole-Rats: New Insights from Patterns of Genetic Relatedness in the Damaraland Mole-Rat (*Cryptomys damarensis*). 269:1025-1030.

Cameron, R. C., E. J. Duncan, and P. K. Dearden. 2013. Biased gene expression in early honeybee larval development. BMC Genom. 14:903.

Cardinal, S. and B. N. Danforth. 2013. Bees diversified in the age of eudicots. 280.

Cardoso-Moreira, M., J. R. Arguello, S. Gottipati, L. G. Harshman, J. K. Grenier, and A. G. Clark. 2016. Evidence for the fixation of gene duplications by positive selection in Drosophila. Genome Res. 26:787-798.

Casneuf, T., S. De Bodt, J. Raes, S. Maere, and Y. Van de Peer. 2006. Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. Genome Biol. 7:R13.

Chapman, R. E. and A. F. G. Bourke. 2001. The influence of sociality on the conservation biology of social insects. Ecol. Lett. 4:650-662.

Chau, L. M. and M. A. D. Goodisman. Gene duplication and the evolution of phenotypic diversity in insect societies.(in press).

Chau, L. M., C. Hanna, L. T. Jenkins, R. E. Kutner, E. A. Burns, C. Kremen, and M. A. D. Goodisman. 2015. Population genetic structure of the predatory, social wasp *Vespula pensylvanica* in its native and invasive range. Ecol. Evol 5:5573-5587.

Ciszek, D. 2000. New colony formation in the "highly inbred" eusocial naked mole-rat: outbreeding is preferred. Behav. Ecol. 11:1-6.

Connallon, T. and A. G. Clark. 2011. The resolution of sexual antagonism by gene duplication. Genetics 187:919-937.

Connallon, T. and L. L. Knowles. 2005. Intergenomic conflict revealed by patterns of sex-biased gene expression. Trends Genet. 21:495-499.

Cornuet, J.-M., P. Pudlo, J. Veyssier, A. Dehne-Garcia, M. Gautier, R. Leblois, J.-M. Marin, and A. Estoup. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data.

Cornuet, J. M. and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001-2014.

Crespi, B. J. 1992. Eusociality in Australian gall thrips. Nature 359:724-726.

Crozier RH, P. P. 1996. Evolution of social insect colonies: Sex allocation and kin selection. Oxford University Press, Oxford.

Cutter, A. D. and S. Ward. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. Mol. Biol. Evol. 22:178-188.

Daly, D., M. E. Archer, P. C. Watts, M. P. Speed, M. R. Hughes, F. S. Barker, J. Jones, K. Odgaard, and S. J. Kemp. 2002. Polymorphic microsatellite loci for eusocial wasps (Hymenoptera: Vespidae). 2:273-275.

Darwin, C. D. 1859. On the origin of Species. Harvard Univ. Press, Cambridge.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. 14:927-930.

Dlugosch, K. M. and I. M. Parker. 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. Mol. Ecol. 17:431-449.

Drake, D. R., C. P. H. Mulder, D. R. Towns, and C. H. Daugherty. 2002. The biology of insularity: an introduction. 29:563-569.

Duffy, J. E. 1996. Eusociality in a coral-reef shrimp. 381:512.

Earl, D. A. and B. M. Vonholdt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359-361.

Earnhardt, J. M., S. D. Thompson, and E. A. Marhevsky. 2001. Interactions of target population size, population parameters, and program management on viability of captive populations. 20:169-183.

Egizi, A., J. Kiser, C. Abadam, and D. M. Fonseca. 2016. The hitchhiker's guide to becoming invasive: exotic mosquitoes spread across a US state by human transport not autonomous flight. 25:3033-3047.

Elango, N., B. G. Hunt, M. A. D. Goodisman, and S. V. Yi. 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. Proc. Natl. Acad. Sci. U. S. A. 106:11206-11211.

Ellegren, H. and J. Parsch. 2007. The evolution of sex-biased genes and sex-biased gene expression. Nat. Rev. Genet. 8:689-698.

Elliott, G. P., P. R. Wilson, R. H. Taylor, and J. R. Beggs. 2010. Declines in common, widespread native birds in a mature temperate forest. Biol. Conserv. 143:2119-2126.

Elsik, C., K. Worley, A. Bennett, M. Beye, F. Camara, C. Childers, D. de Graaf, G. Debyser, J. Deng, B. Devreese, E. Elhaik, J. Evans, L. Foster, D. Graur, R. Guigo, H. p. teams, K. Hoff, M. Holder, M. Hudson, G. Hunt, H. Jiang, V. Joshi, R. Khetani, P. Kosarev, C. Kovar, J. Ma, R. Maleszka, R. Moritz, M. Munoz-Torres, and T. Murphy. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade. BMC Genom. 15:86.

Emlen, S. T., J. M. Emlen, and S. A. Levin. 1986. Sex-ratio selection in species with helpers-at-the-nest. 127:1-8.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. 14:2611-2620.

Evans, T. A., B. T. Forschler, and J. K. Grace. 2013. Biology of Invasive Termites: A Worldwide Review. Annu. Rev. Entomol. 58:455-474.

Fares, M. A., O. M. Keane, C. Toft, L. Carretero-Paulet, and G. W. Jones. 2013. The roles of whole-genome and small-scale duplications in the functional specialization of *Saccharomyces cerevisiae* genes. Pp. e1003176. PLoS Genet. Public Library of Science.

Farré, D. and M. M. Albà. 2010. Heterogeneous patterns of gene-expression diversification in mammalian gene duplicates. Mol. Biol. Evol. 27:325-335.

Faulkes, C. G., D. H. Abbott, H. P. O'Brien, L. Lau, M. R. Roy, R. K. Wayne, and M. W. Bruford. 1997a. Micro- and macrogeographical genetic structure of colonies of naked mole-rats *Heterocephalus glaber*. 6:615-628.

Faulkes, C. G. and N. C. Bennett. 2001. Family values: group dynamics and social control of reproduction in African mole-rats. Trends Ecol. Evol. 16:184-190.

Faulkes, C. G. and N. C. Bennett. 2013. Plasticity and constraints on social evolution in African mole-rats: ultimate and proximate factors. 368.

Faulkes, C. G., N. C. Bennett, M. W. Bruford, H. P. O'Brien, G. H. Aguilar, and J. U. M. Jarvis. 1997b. Ecological constraints drive social evolution in the African mole-rats. 264:1619-1627.

Feldmeyer, B., D. Elsner, and S. Foitzik. 2014. Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. Mol. Ecol. 23:151-161.

Felsenstein, J. 1985. Phylogenies and the comparative method. 125:1-15.

Ferreira, P. G., S. Patalano, R. Chauhan, R. Ffrench-Constant, T. Gabaldon, R. Guigo, and S. Sumner. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. Genome Biol. 14.

Fisher, R. A. 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Fonzi, E., Y. Higa, A. G. Bertuso, K. Futami, and N. Minakawa. 2015. Human-mediated marine dispersal influences the population structure of *Aedes aegypti* in the Philippine archipelago. 9:e0003829.

Foote, D. and H. L. Carson. 1995. Drosophila as monitors of change in Hawaiian ecosystems. Pp. 368-372. Our living resources: a report to the nation on the distribution, abundance, and health of U.S. plants, animals and ecosystems. U.S. Department of the Interior, National Biological Service.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531-1545.

Foret, S., R. Kucharski, Y. Pittelkow, G. Lockett, and R. Maleszka. 2009. Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. 10:472.

Frankham, R. 2008. Genetic adaptation to captivity in species conservation programs. 17:325-333.

Freilich, S., T. Massingham, E. Blanc, L. Goldovsky, and J. M. Thornton. 2006. Relating tissue specialization to the differentiation of expression of singleton and duplicate mouse proteins. Genome Biol. 7:R89.

Gadagkar, R. 1990. Evolution of eusociality - The advantage of assured fitness returns. Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci. 329:17-25.

Gadagkar, R. 1991. Demographic predisposition to the evolution of eusociality - A hierarchy of models. Proc. Natl. Acad. Sci. U. S. A. 88:10993-10997.

Gadagkar, R. 1997a. The evolution of caste polymorphism in social insects: Genetic release followed by diversifying evolution. 76:167-179.

Gadagkar, R. 1997b. The evolution of caste polymorphism in social insects: genetic release followed by diversifying evolution. J. Genet. 76:167-179.

Gallach, M. and E. Betran. 2011. Intralocus sexual conflict resolved through gene duplication. Trends Ecol. Evol. 26:222-228.

Galtier, N. 2016. Adaptive protein evolution in animals and the effective population size hypothesis. Pp. e1005774. PLoS Genet. Public Library of Science, San Francisco, CA USA.

Gambino, P. 1991. Reproductive plasticity of *Vespula pensylvanica* (Hymenoptera, Vespidae) on Maui and Hawaii Islands, USA. N. Z. J. Zool. 18:139-149.

Gambino, P. and L. Loope. 1992. Yellowjacket (*Vespula pensylvanica*) biology and abatement in the National Parks of Hawaii.i-v, 1-41.

Goodisman, M. A. D., R. W. Matthews, and R. H. Crozier. 2001. Hierarchical genetic structure of the introduced wasp *Vespula germanica* in Australia. Mol. Ecol. 10:1423-1432.

Gotzek, D., H. J. Axen, A. V. Suarez, S. H. Cahan, and D. Shoemaker. 2015. Global invasion history of the tropical fire ant: a stowaway on the first global trade routes. Mol. Ecol. 24:374-388.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. J. Hered. 86:485-486.

Gowaty, P. A. and M. R. Lennartz. 1985. Sex-raatios of nestling and fledinging Red-cockaded woodpeckers (*Picoides borealis*) favor males. Am. Nat. 126:347-353.

Gu, X. 2003. Evolution of duplicate genes versus genetic robustness against null mutations. 19.

Hall, D. W., S. V. Yi, and M. A. D. Goodisman. 2013. Kin selection, genomics and caste-antagonistic pleiotropy. Biol. Lett. 9.

Hamilton, W. D. 1964. The genetical evolution of social behavior. 7:1-16.

Hanna, C., E. D. Cook, A. R. Thompson, L. E. Dare, A. L. Palaski, D. Foote, and M. A. D. Goodisman. 2014a. Colony social structure in native and invasive populations of the social wasp *Vespula pensylvanica*. Biol. Invasions 16:283-294.

Hanna, C., D. Foote, and C. Kremen. 2014b. Competitive impacts of an invasive nectar thief on plant-pollinator mutualisms. Ecology 95:1622-1632.

Harpur, B. A., C. F. Kent, D. Molodtsova, J. M. D. Lebon, A. S. Alqarni, A. A. Owayss, and A. Zayed. 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits.

Harrison, M. C., R. L. Hammond, and E. B. Mallon. 2015. Reproductive workers show queenlike gene expression in an intermediately eusocial insect, the buff-tailed bumble bee *Bombus terrestris*. Mol. Ecol. 24:3043-3063.

Hasegawa, E. and J. Takahashi. 2002. Microsatellite loci for genetic research in the hornet Vespa mandarinia and related species. 2:306-308.

He, X. L. and J. Z. Zhang. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157-1164.

Helantera, H., J. E. Strassmann, J. Carrillo, and D. C. Queller. 2009. Unicolonial ants: where do they come from, what are they and where are they going? Trends Ecol. Evol. 24:341-349.

Herb, B. R., F. Wolschin, K. D. Hansen, M. J. Aryee, B. Langmead, R. Irizarry, G. V. Amdam, and A. P. Feinberg. 2012. Reversible switching between epigenetic states in honeybee behavioral subcastes. Nat. Neurosci. 15:1371-1373.

Hoffman, E. A. and M. A. D. Goodisman. 2007. Gene expression and the evolution of phenotypic diversity in social wasps. BMC Biol. 5.

Hoffman, E. A., J. L. Kovacs, and M. A. D. Goodisman. 2008. Genetic structure and breeding system in a social wasp and its social parasite. BMC Evol. Biol. 8.

Holldobler, E. and E. O. Wilson. 1990. The Ants. Belknap Press of Harvard University Press.

Holman, L. 2014. Caste load and the evolution of reproductive skew. Am. Nat. 183:84-95.

Holsinger, K. E. and B. S. Weir. 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat. Rev. Genet. 10:639-650.

Holway, D. A., L. Lach, A. V. Suarez, N. D. Tsutsui, and T. J. Case. 2002. The causes and consequences of ant invasions. 33:181-233.

Huminiecki, L. and K. H. Wolfe. 2004. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. Genome Res. 14:1870-1879.

Hunt, B. G., L. Ometto, L. Keller, and M. A. D. Goodisman. 2013. Evolution at two levels in fire ants: The relationship between patterns of gene expression and protein sequence evolution. Mol. Biol. Evol. 30:263-271.

Hunt, J. H., N. S. Mutti, H. Havukainen, M. T. Henshaw, and G. V. Amdam. 2011. Development of an RNA Interference tool, characterization of its target, and an ecological test of caste differentiation in the eusocial wasp Polistes. PLoS One 6.

Husseneder, C., D. M. Simms, J. R. Delatte, C. L. Wang, J. K. Grace, and E. L. Vargo. 2012. Genetic diversity and colony breeding structure in native and introduced ranges of the Formosan subterranean termite, *Coptotermes formosanus*. Biol. Invasions 14:419-437.

Ibn-Salem, J., E. M. Muro, and M. A. Andrade-Navarro. 2016. Co-regulation of paralog genes in the three-dimensional chromatin architecture. Nucleic Acids Res.

Ingram, C. M., N. J. Troendle, C. A. Gill, S. Braude, and R. L. Honeycutt. 2015. Challenging the inbreeding hypothesis in a eusocial mammal: population genetics of the naked mole-rat, *Heterocephalus glaber*. 24:4848-4865.

Ingram, C. M., N. J. Troendle, C. A. Gill, and R. L. Honeycutt. 2014. Development of 12 new microsatellite markers for the naked mole-rat, *Heterocephalus glaber*. 6:589-591.

Innan, H. and F. Kondrashov. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11:97-108.

Innocenti, P. and E. H. Morrow. 2010. The sexually antagonistic genes of *Drosophila melanogaster*. PLoS. Biol.

Ivy, J. A. and R. C. Lacy. 2012. A Comparison of Strategies for Selecting Breeding Pairs to Maximize Genetic Diversity Retention in Managed Populations. J. Hered. 103:186-196.

Jakobsson, M. and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806.

Jarvis, J. U. M. 1981. Eusociality in a mammal - cooperative breeding in naked mole rat colonies. Science 212:571-573.

Jarvis, J. U. M. and N. C. Bennett. 1993. Eusociality has evolved independently in two genera of bathyergid mole-rats — but occurs in no other subterranean mammal. 33:253-260.

Jarvis, J. U. M., M. J. Oriain, N. C. Bennett, and P. W. Sherman. 1994. Mammalian eusociality - A family affair. Trends Ecol. Evol. 9:47-51.

Jasper, W. C., T. A. Linksvayer, J. Atallah, D. Friedman, J. C. Chiu, and B. R. Johnson. 2015. Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees. Mol. Biol. Evol. 32:334-346.

Jeanne, R. L. 1986. The organization of work in *Polybia occidentalis*: costs and benefits of specialization in a social wasp. 19:333-341.

Jensen, J. D. and D. Bachtrog. 2011. Characterizing the influence of effective population size on the rate of adaptation: Gillespie's Darwin domain. Genome Biol. Evol. 3:687-701.

Johnson, B. R. and N. D. Tsutsui. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. BMC Genom. 12:164.

Kapheim, K. M., H. Pan, C. Li, S. L. Salzberg, D. Puiu, T. Magoc, H. M. Robertson, M. E. Hudson, A. Venkat, B. J. Fischman, A. Hernandez, M. Yandell, D. Ence, C. Holt, G. D. Yocum, W. P. Kemp, J. Bosch, R. M. Waterhouse, E. M. Zdobnov, E. Stolle, F. B. Kraus, S. Helbing, R. F. A. Moritz, K. M. Glastad, B. G. Hunt, M. A. D. Goodisman, F. Hauser, C. J. P. Grimmelikhuijzen, D. G. Pinheiro, F. M. F. Nunes, M. P. M. Soares, É. D. Tanaka, Z. L. P. Simões, K. Hartfelder, J. D. Evans, S. M. Barribeau, R. M. Johnson, J. H. Massey, B. R. Southey, M. Hasselmann, D. Hamacher, M. Biewer, C. F. Kent, A. Zayed, C. Blatti, S. Sinha, J. S. Johnston, S. J. Hanrahan, S. D. Kocher, J. Wang, G. E. Robinson, and G. Zhang. 2015. Genomic signatures of evolutionary transitions from solitary to group living. 348:1139-1143.

Katsushima, K., C. Nishida, S. Yosida, M. Kato, K. Okanoya, and Y. Matsuda. 2010. A multiplex PCR assay for molecular sexing of the naked mole-rat (*Heterocephalus glaber*). Mol. Ecol. Resour. 10:222-224.

Kawalika, M. and H. Burda. 2007. Giant mole-rats, *Fukomys mechowii*, 13 years on the stage.205-219.

Kenis, M., M. A. Auger-Rozenberg, A. Roques, L. Timms, C. Pere, M. Cock, J. Settele, S. Augustin, and C. Lopez-Vaamonde. 2009. Ecological effects of invasive alien insects. Biol. Invasions 11:21-45.

Kent, D. S. and J. A. Simpson. 1992. Eusociality in the beetle *Austroplatypus incompertus* (Coleoptera: Curculionidae). 79:86-87.

Kim, E. B., X. D. Fang, A. A. Fushan, Z. Y. Huang, A. V. Lobanov, L. J. Han, S. M. Marino, X. Q. Sun, A. A. Turanov, P. C. Yang, S. H. Yim, X. Zhao, M. V. Kasaikina, N. Stoletzki, C. F. Peng, P. Polak, Z. Q. Xiong, A. Kiezun, Y. B. Zhu, Y. X. Chen, G. V. Kryukov, Q. Zhang, L. Peshkin, L. Yang, R. T. Bronson, R. Buffenstein, B. Wang, C. L. Han, Q. Y. Li, L. Chen, W. Zhao, S. R. Sunyaev, T. J. Park, G. J. Zhang, J. Wang, and V. N. Gladyshev. 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature 479:223-227.

Kirk, H., S. Dorn, and D. Mazzi. 2013. Molecular genetics and genomics generate new insights into invertebrate pest invasions. Evol. Appl. 6:842-856.

Kocher, S., C. Li, W. Yang, H. Tan, S. Yi, X. Yang, H. Hoekstra, G. Zhang, N. Pierce, and D. Yu. 2013. The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. 14:R142.

Kolbe, J. J., R. E. Glor, L. R. G. Schettino, A. C. Lara, A. Larson, and J. B. Losos. 2004. Genetic variation increases during biological invasion by a Cuban lizard. Nature 431:177-181.

Kovacs, J. L., E. A. Hoffman, S. M. Marriner, and M. A. D. Goodisman. 2010. Detecting selection on morphological traits in social insect castes: the case of the social wasp *Vespula maculifrons*. Biol. J. Linnean Soc. 101:93-102.

Kucharski, R., J. Maleszka, S. Foret, and R. Maleszka. 2008. Nutritional control of reproductive status in honeybees via DNA methylation. Science 319:1827-1830.

Lacey, E. A. and P. W. Sherman. 1991. Social organization of naked mole-rat (*Heterocephalus glaber*) colonies: evidence for a division of labour. Pp. 275-336 *in* P. W. Sherman, J. U. M. Jarvis, and R. D. Alexander, eds. The biology of the naked mole-rat. Princeton University Press, Princeton, NJ.

Lacy, R. C. 2013. Achieving True Sustainability of Zoo Populations. Zoo Biol. 32:19-26.

Lan, X. and J. K. Pritchard. 2016. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. Science 352:1009-1013.

Lande, R. 1980. Sexual Dimorphism, Sexual Selection, and Adaptation in Polygenic Characters. 34:292-305.

Landolt, P. J., H. C. Reed, and D. J. Ellis. 2003. Trapping yellowjackets (Hymenoptera : Vespidae) with heptyl butyrate emitted from controlled-release dispensers. Fla. Entomol. 86:323-328.

Langmead, B. and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. 9:357-U354.

Le Comber, S. C., A. C. Spinks, N. C. Bennett, J. U. M. Jarvis, and C. G. Faulkes. 2002. Fractal dimension of African mole-rat burrows. Can. J. Zool.-Rev. Can. Zool. 80:436-441.

Leichty, A. R., D. W. Pfennig, C. D. Jones, and K. S. Pfennig. 2012. Relaxed genetic constraint is ancestral to the evolution of phenotypic plasticity. Integr. Comp. Biol. 52:16-30.

Li, B. and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. 12:323.

Lin, N. and C. D. Michener. 1972. Evolution of Sociality in Insects. 47:131-159.

Linksvayer, T. A. and M. J. Wade. 2005. The evolutionary origin and elaboration of sociality in the aculeate hymenoptera: maternal effects, sib-social effects, and heterochrony. Q. Rev. Biol. 80:317-336.

Lockett, G. A., E. J. Almond, T. J. Huggins, J. D. Parker, and A. F. G. Bourke. 2016. Gene expression differences in relation to age and social environment in queen and worker bumble bees. Exp. Gerontol. 77:52-61.

Luikart, G., F. Allendorf, J.-M. Cornuet, and W. Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. 89:238-247.

Lutermann, H., E. Z. Cameron, N. J. Raihani, and N. C. Bennett. 2014. Sex ratio variation in a eusocial mammal, the Damaraland mole-rat, *Fukomys damarensis*. 294:139-145.

Lyko, F. and R. Maleszka. 2011. Insects as innovative models for functional studies of DNA methylation. Trends Genet. 27:127-131.

Lynch, M. and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155.

Mank, J. E. and H. Ellegren. 2009. Are sex-biased genes more dispensable? Biol. Lett. 5:409-412.

Masciocchi, M. and J. Corley. 2013. Distribution, dispersal and spread of the invasive social wasp (*Vespula germanica*) in Argentina. Austral Ecol. 38:162-168.

Matthews, R. W., M. A. D. Goodisman, A. D. Austin, and R. Bashford. 2000. The introduced English wasp *Vespula vulgaris* (L.) (Hymenoptera : Vespidae) newly recorded invading native forests in Tasmania. Aust. J. Entomol. 39:177-179.

Michener, C. D. 1969. Comparative Social Behavior of Bees. 14:299-342.

Mikhaylova, L. M., K. Nguyen, and D. I. Nurminsky. 2008. Analysis of the *Drosophila melanogaster* testes transcriptome reveals coordinate regulation of paralogous genes. 179:305-315.

Moller, H. 1996. Lessons for invasion theory from social insects. Biol. Conserv. 78:125-142.

Monceau, K., O. Bonnard, and D. Thiery. 2014. *Vespa velutin*a: a new invasive predator of honeybees in Europe. J. Pest Sci. 87:1-16.

Mooney, S. J., D. C. S. Filice, N. R. Douglas, and M. M. Holmes. 2015. Task specialization and task switching in eusocial mammals. 109:227-233.

Moran, E. V. and J. M. Alexander. 2014. Evolutionary responses to global change: lessons from invasive species. Ecol. Lett. 17:637-649.

Morandin, C., K. Dhaygude, J. Paviala, K. Trontti, C. Wheat, and H. Helantera. 2015. Caste-biases in gene expression are specific to developmental stage in the ant *Formica exsecta*. J. Evol. Biol. 28:1705-1718.

Muirhead, J. R., D. K. Gray, D. W. Kelly, S. M. Ellis, D. D. Heath, and H. J. Macisaac. 2008. Identifying the source of species invasions: sampling intensity vs. genetic diversity. Mol. Ecol. 17:1020-1035.

Nakahara, L. M. 1980. Survey report on the Yellowjackets, *Vespula pensylvanica* and *Vespula vulgaris*, in Hawaii *in* H. D. o. Agriculture, ed.

Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. J. Mol. Evol. 19:153-170.

Normark, B. B. 2003. The evolution of alternative genetic systems in insects. Annu. Rev. Entomol. 48:397-423.

Nygaard, S., G. Zhang, M. Schiøtt, C. Li, Y. Wurm, H. Hu, J. Zhou, L. Ji, F. Qiu, M. Rasmussen, H. Pan, F. Hauser, A. Krogh, C. J. P. Grimmelikhuijzen, J. Wang, and

J. J. Boomsma. 2011. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming.

Ohno, S. 1970. Evolution by Gene Duplication. Springer.

Ometto, L., D. Shoemaker, K. G. Ross, and L. Keller. 2011. Evolution of gene expression in fire ants: The effects of developmental stage, caste, and species. 28:1381-1392.

Oriain, M. J., J. U. M. Jarvis, and C. G. Faulkes. 1996. A dispersive morph in the naked mole-rat. Nature 380:619-621.

Oxley, Peter R., L. Ji, I. Fetter-Pruneda, Sean K. McKenzie, C. Li, H. Hu, G. Zhang, and Daniel J. C. Kronauer. 2014. The Genome of the Clonal Raider Ant *Cerapachys biroi*. 24:451-458.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. 20:289-290.

Patalano, S., A. Vlasova, C. Wyatt, P. Ewels, F. Camara, P. G. Ferreirab, C. L. Asher, T. P. Jurkowski, A. Segonds-Pichon, M. Bachman, I. Gonzalez-Navarrete, A. E. Minoche, F. Krueger, E. Lowy, M. Marcet-Houben, J. L. Rodriguez-Ales, F. S. Nascimento, S. Balasubramanian, T. Gabaldon, J. E. Tarver, S. Andrews, H. Himmelbauer, W. O. H. Hughes, R. Guigo, W. Reik, and S. Sumner. 2015. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. Proc. Natl. Acad. Sci. U. S. A. 112:13970-13975.

Peakall, R. and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. 28:2537-2539.

Pejchar, L. and H. A. Mooney. 2009. Invasive species, ecosystem services and human well-being. Trends Ecol. Evol. 24:497-504.

Pemberton, J. 2004. Measuring inbreeding depression in the wild: the old ways are the best. Trends Ecol. Evol. 19:613-615.

Pennell, T. M. and E. H. Morrow. 2013. Two sexes, one genome: the evolutionary dynamics of intralocus sexual conflict. Ecol. Evol. 3:1819-1834.

Pereboom, J. J. M., W. C. Jordan, S. Sumner, R. L. Hammond, and A. F. G. Bourke. 2005. Differential gene expression in queen-worker caste determination in bumble-bees. Proc. R. Soc. Lond. Ser. B-Biol. Sci. 272:1145-1152.

Perry, B. R. and R. Assis. 2016. CDROM: Classification of Duplicate gene RetentiOn Mechanisms. BMC Evol. Biol. 16:1-4.

Pierce, A. A., M. P. Zalucki, M. Bangura, M. Udawatta, M. R. Kronforst, S. Altizer, J. F. Haeger, and J. C. de Roode. 2014. Serial founder effects and genetic differentiation during worldwide range expansion of monarch butterflies. 281.

Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. J. Hered. 95:536-539.

Piry, S., G. Luikart, and J. Cornuet. 1999. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. 90:502-503.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. 155:945-959.

Proulx, S. R. 2012. Multiple routes to subfunctionalization and gene duplicate specialization. 190:737-751.

Purcell, K. M., N. Ling, and C. A. Stockwell. 2012. Evaluation of the introduction history and genetic diversity of a serially introduced fish population in New Zealand. Biol. Invasions 14:2057-2065.

Queller, D. C. and J. E. Strassmann. 1998. Kin Selection and Social Insects. 48:165-175.

Ranwez, V., S. Harispe, F. Delsuc, and E. J. P. Douzery. 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. Pp. e22594. PLoS One. Public Library of Science.

Reeve, H. K., D. F. Westneat, W. A. Noon, P. W. Sherman, and C. F. Aquadro. 1990. DNA fingerprinting reveals high-levels of inbreeding in colonies of the eusocial Naked mole-rat. Proc. Natl. Acad. Sci. U. S. A. 87:2496-2500.

Rehan, S. M., R. Leys, and M. P. Schwarz. 2012. A Mid-Cretaceous Origin of Sociality in Xylocopine Bees with Only Two Origins of True Worker Castes Indicates Severe Barriers to Eusociality. 7:e34690.

Rehan, S. M., M. P. Schwarz, and M. H. Richards. 2011. Fitness consequences of ecological constraints and implications for the evolution of sociality in an incipiently social bee. 103:57-67.

Rehan, S. M. and A. L. Toth. 2015. Climbing the social ladder: the molecular evolution of sociality. 30:426-433.

Rice, W. R. 2013. Nothing in genetics makes sense except in light of genomic conflict. Pp. 217-237 *in* D. J. Futuyma, ed. Annual Review of Ecology Evolution and Systematics. Annual Reviews, Palo Alto.

Rice, W. R. and A. K. Chippindale. 2001. Intersexual ontogenetic conflict. J. Evol. Biol. 14:685-693.

Robinson, G. E., C. M. Grozinger, and C. W. Whitfield. 2005. Sociogenomics: social life in molecular terms. Nature reviews. Genetics 6:257-270.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140.

Roderick, G. K. and R. G. Gillespie. 1998. Speciation and phylogeography of Hawaiian terrestrial arthropods. 7:519-531.

Romiguier, J., J. Lourenco, P. Gayral, N. Faivre, L. A. Weinert, S. Ravel, M. Ballenghien, V. Cahais, A. Bernard, E. Loire, L. Keller, and N. Galtier. 2014. Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. J. Evol. Biol. 27:593-603.

Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes 4:137-138.

Rousset, F. 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. 8:103-106.

Roux, J., E. Privman, S. Moretti, J. T. Daub, M. Robinson-Rechavi, and L. Keller. 2014. Patterns of Positive Selection in Seven Ant Genomes. 31:1661-1685.

Sadd, B. M., S. M. Barribeau, G. Bloch, D. C. de Graaf, P. Dearden, C. G. Elsik, J. Gadau, C. J. P. Grimmelikhuijzen, M. Hasselmann, J. D. Lozier, H. M. Robertson, G. Smagghe, E. Stolle, M. Van Vaerenbergh, R. M. Waterhouse, E. Bornberg-Bauer, S. Klasberg, A. K. Bennett, F. Caamara, R. Guigo, K. Hoff, M. Mariotti, M. Munoz-Torres, T. Murphy, D. Santesmasses, G. V. Amdam, M. Beckers, M. Beye, M. Biewer, M. M. G. Bitondi, M. L. Blaxter, A. F. G. Bourke, M. J. F. Brown, S. D. Buechel, R. Cameron, K. Cappelle, J. C. Carolan, O. Christiaens, K. L. Ciborowski, D. F. Clarke, T. J. Colgan, D. H. Collins, A. G. Cridge, T. Dalmay, S. Dreier, L. du Plessis, E. Duncan, S. Erler, J. Evans, T. Falcon, K. Flores, F. C. P. Freitas, T. Fuchikawa, T. Gempe, K. Hartfelder, F. Hauser, S. Helbing, F. C. Humann, F. Irvine, L. S. Jermiin, C. E. Johnson, R. M. Johnson, A. K. Jones, T. Kadowaki, J. H. Kidner, V. Koch, A. Kohler, F. B. Kraus, H. M. G. Lattorff, M. Leask, G. A. Lockett, E. B. Mallon, D. S. M. Antonio, M. Marxer, I. Meeus, R. F. A. Moritz, A. Nair, K. Napflin, I. Nissen, J. Niu, F. M. F. Nunes, J. G. Oakeshott, A. Osborne, M. Otte, D. G. Pinheiro, N. Rossie, O. Rueppell, C. G. Santos, R. Schmid-Hempel, B. D. Schmitt, C. Schulte, Z. L. P. Simoes, M. P. M. Soares, L. Swevers, E. C. Winnebeck, F. Wolschin, N. Yu, E. M. Zdobnov, P. K. Aqrawi, K. P. Blankenburg, M. Coyle, L. Francisco, A. G. Hernandez, M. Holder, M. E. Hudson, L. Jackson, J. Jayaseelan, V. Joshi, C. Kovar, S. L. Lee, R. Mata, T. Mathew, I. F. Newsham, R. Ngo, G. Okwuonu, C. Pham, L. L. Pu, N. Saada, J. Santibanez, D. Simmons, R. Thornton, A. Venkat, K. K. O. Walden, Y. Q. Wu, G. Debyser, B. Devreese, C. Asher, J. Blommaert, A. D. Chipman, L. Chittka, B. Fouks, J. Liu, M. P. O'Neill, S. Sumner, D. Puiu, J. Qu, S. L. Salzberg, S. E. Scherer, D. M. Muzny, S. Richards, G. E. Robinson, R. A. Gibbs, P. Schmid-Hempel and K. C. Worley. 2015. The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. 16.

Sakai, A. K., F. W. Allendorf, J. S. Holt, D. M. Lodge, J. Molofsky, K. A. With, S. Baughman, R. J. Cabin, J. E. Cohen, N. C. Ellstrand, D. E. McCauley, P. O'Neil, I. M. Parker, J. N. Thompson, and S. G. Weller. 2001. The population biology of invasive species. Annu. Rev. Ecol. Syst. 32:305-332.

Schmid-Hempel, P., R. Schmid-Hempel, P. C. Brunner, O. D. Seeman, and G. R. Allen. 2007. Invasion success of the bumblebee, *Bombus terrestris*, despite a drastic genetic bottleneck. 99:414-422.

Schmid-Hempel, R., M. Eckhardt, D. Goulson, D. Heinzmann, C. Lange, S. Plischuk, L. R. Escudero, R. Salathe, J. J. Scriven, and P. Schmid-Hempel. 2014. The invasion of southern South America by imported bumblebees and associated parasites. J. Anim. Ecol. 83:823-837.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. 18:233-234.

Seeley, T. D. 1982. Adaptive significance of the age polyethism schedule in honeybee colonies. Behav. Ecol. Sociobiol. 11:287-293.

Sherman, P. W., J. U. M. Jarvis, and S. H. Braude. 1992. NAKED MOLE RATS. Sci.Am. 267:72-78.

Silk, J. B. 1983. Local Resource Competition and Facultative Adjustment of Sex Ratios in Relation to Competitive Abilities. 121:56-66.

Simberloff, D., J. L. Martin, P. Genovesi, V. Maris, D. A. Wardle, J. Aronson, F. Courchamp, B. Galil, E. Garcia-Berthou, M. Pascal, P. Pysek, R. Sousa, E. Tabacchi, and M. Vila. 2013. Impacts of biological invasions: what's what and the way forward. Trends Ecol. Evol. 28:58-66.

Smith, C. D., A. Zimin, C. Holt, E. Abouheif, R. Benton, E. Cash, V. Croset, C. R. Currie, E. Elhaik, C. G. Elsik, M. J. Fave, V. Fernandes, J. Gadau, J. D. Gibson, D. Graur, K. J. Grubbs, D. E. Hagen, M. Helmkampf, J. A. Holley, H. Hu, A. S. I. Viniegra, B. R. Johnson, R. M. Johnson, A. Khila, J. W. Kim, J. Laird, K. A. Mathis, J. A. Moeller, M. C. Munoz-Torres, M. C. Murphy, R. Nakamura, S. Nigam, R. P. Overson, J. E. Placek, R. Rajakumar, J. T. Reese, H. M. Robertson, C. R. Smith, A. V. Suarez, G. Suen, E. L. Suhr, S. Tao, C. W. Torres, E. van Wilgenburg, L. Viljakainen, K. K. O. Walden, A. L. Wild, M. Yandell, J. A. Yorke, and N. D. Tsutsui. 2011. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). Proc. Natl. Acad. Sci. U. S. A. 108:5673-5678.

Smith, C. R., A. L. Toth, A. V. Suarez, and G. E. Robinson. 2008. Genetic and genomic analyses of the division of labour in insect societies. Nature reviews. Genetics 9:735-748.

Standage, D. S., A. J. Berens, K. M. Glastad, A. J. Severin, V. P. Brendel, and A. L. Toth. 2016. Genome, transcriptome and methylome sequencing of a primitively eusocial

wasp reveal a greatly reduced DNA methylation system in a social insect. Mol. Ecol. 25:1769-1784.

Stern, D. L. and W. A. Foster. 1996. The evolution of soldiers in aphids. 71:27-79.

Stewart, A. D., A. Pischedda, and W. R. Rice. 2010. Resolving intralocus sexual conflict: genetic mechanisms and time frame. J. Hered. 101:S94-S99.

Suarez, A. V. and T. J. Case. 2002. Bottom-up effects on persistence of a specialist predator : Ant invasions and horned lizards. Ecol. Appl. 12:291-298.

Suarez, A. V. and N. D. Tsutsui. 2008. The evolutionary consequences of biological invasions. 17:351-360.

Suen, G., C. Teiling, L. Li, C. Holt, E. Abouheif, E. Bornberg-Bauer, P. Bouffard, E. J. Caldera, E. Cash, A. Cavanaugh, O. Denas, E. Elhaik, M. J. Fave, J. R. Gadau, J. D. Gibson, D. Graur, K. J. Grubbs, D. E. Hagen, T. T. Harkins, M. Helmkampf, H. Hu, B. R. Johnson, J. Kim, S. E. Marsh, J. A. Moeller, M. C. Munoz-Torres, M. C. Murphy, M. C. Naughton, S. Nigam, R. Overson, R. Rajakumar, J. T. Reese, J. J. Scott, C. R. Smith, S. Tao, N. D. Tsutsui, L. Viljakainen, L. Wissler, M. D. Yandell, F. Zimmer, J. Taylor, S. C. Slater, S. W. Clifton, W. C. Warren, C. G. Elsik, C. D. Smith, G. M. Weinstock, N. M. Gerardo, and C. R. Currie. 2011. The Genome Sequence of the Leaf-Cutter Ant *Atta cephalotes* Reveals Insights into Its Obligate Symbiotic Lifestyle. PLoS Genet. 7.

Sumner, S. 2014. The importance of genomic novelty in social evolution. Mol. Ecol. 23:26-28.

Sumner, S., J. J. M. Pereboom, and W. C. Jordan. 2006. Differential gene expression and phenotypic plasticity in behavioural castes of the primitively eusocial wasp, *Polistes canadensis*. Proc. R. Soc. B-Biol. Sci. 273:19-26.

Szathmary, E. and J. M. Smith. 1995. The major evolutionary transitions. Nature 374:227-232.

Takezaki, N., M. Nei, and K. Tamura. 2010. POPTREE2: Software for Constructing Population Trees from Allele Frequency Data and Computing Other Population Statistics with Windows Interface. Mol. Biol. Evol. 27:747-752.

Tautz, D. and T. Domazet-Lošo. 2011. The evolutionary origin of orphan genes. Nat. Rev. Genet. 12:692-702.

Terrapon, N., C. Li, H. M. Robertson, L. Ji, X. H. Meng, W. Booth, Z. S. Chen, C. P. Childers, K. M. Glastad, K. Gokhale, J. Gowin, W. Gronenberg, R. A. Hermansen, H. F. Hu, B. G. Hunt, A. K. Huylmans, S. M. S. Khalil, R. D. Mitchell, M. C. Munoz-Torres, J. A. Mustard, H. L. Pan, J. T. Reese, M. E. Scharf, F. M. Sun, H. Vogel, J. Xiao, W. Yang, Z. K. Yang, Z. Q. Yang, J. J. Zhou, J. W. Zhu, C. S. Brent, C. G. Elsik, M. A. D. Goodisman, D. A. Liberles, R. M. Roe, E. L. Vargo, A.

Vilcinskas, J. Wang, E. Bornberg-Bauer, J. Korb, G. J. Zhang, and J. Liebig. 2014. Molecular traces of alternative social organization in a termite genome. Nat. Commun. 5.

Thiel, T., W. Michalek, R. Varshney, and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet 106:411-422.

Thoren, P. A., R. J. Paxton, and A. Estoup. 1995. Unusually high frequency of (CT)n and (GT)n microsatellite loci in a yellowjacket wasp, *Vespula rufa* (L.) (Hymenoptera: Vespidae). Insect Mol. Biol. 4:141-148.

Toth, A. L., K. Varala, M. T. Henshaw, S. L. Rodriguez-Zas, M. E. Hudson, and G. E. Robinson. 2010. Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. Proc. R. Soc. B-Biol. Sci. 277:2139-2148.

Trivers, R. L. and H. Hare. 1976. Haploidiploidy and evolution of social insects. Science 191:249-263.

Trivers, R. L. and D. E. Willard. 1973. Natural selection of parental ability to vary the sex ratio of offspring. 179:90-92.

True, J. R. and S. B. Carroll. 2002. Gene co-option in physiological and morphological evolution. 18:53-80.

Tsuchida, K., K. Kudo, and N. Ishiguro. 2014. Genetic structure of an introduced paper wasp, *Polistes chinensis antennalis* (Hymenoptera, Vespidae) in New Zealand. 23:4018-4034.

Tsutsui, N. D. and A. V. Suarez. 2003. The colony structure and population biology of invasive ants. Conserv. Biol. 17:48-58.

Tsutsui, N. D., A. V. Suarez, D. A. Holway, and T. J. Case. 2000. Reduced genetic variation and the success of an invasive species. 97:5948-5953.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. Primer3—new capabilities and interfaces. 40:e115-e115.

van Oosterhout, C., D. Weetman, and W. F. Hutchinson. 2006. Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. Mol. Ecol. Notes 6:255-256.

Vargo, E. L. 2003. Hierarchical analysis of colony and population genetic structure of the eastern subterranean termite, *Reticulitermes flavipes*, using two classes of molecular markers. Evolution; international journal of organic evolution 57:2805-2818.

Visscher, P. K. and R. S. Vetter. 2003. Annual and multi-year nests of the western yellowjacket, *Vespula pensylvanica*, in California. Insect. Soc. 50:160-166.

Vleurinck, C., S. Raub, D. Sturgill, B. Oliver, and M. Beye. 2016. Linking genes and brain development of honeybee workers: A whole-transcriptome approach. Pp. e0157980. PLoS One. Public Library of Science.

Wang, J., F. Tao, N. C. Marowsky, and C. Fan. 2016. Evolutionary fates and dynamic functionalization of young duplicate genes in Arabidopsis genomes. Plant Physiol.

Wcislo, W. T. and J. H. Cane. 1996. Floral resource utilization by solitary bees (Hymenoptera: Apoidea) and exploitation of their stored foods by natural enemies. 41:257-286.

Weir, B. S. 1996. Genetic Data Analysis. Sinauer, Sunderland, MA.

Weir, B. S. and C. C. Cockerham. 1984. F-statistcs for the analysis of population structure. Evolution; international journal of organic evolution 38:1358-1370.

West-Eberhard, M. J. 1989. Phenotypic Plasticity and the Origins of Diversity. 20:249-278.

West, S. A., R. M. Fisher, A. Gardner, and E. T. Kiers. 2015. Major evolutionary transitions in individuality. 112:10112-10119.

West, S. A., A. S. Griffin, and A. Gardner. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. 20:415-432.

Whitfield, C. W., A. M. Cziko, and G. E. Robinson. 2003. Gene expression profiles in the brain predict behavior in individual honey bees. Science 302:296-299.

Wilson, E. E. and D. A. Holway. 2010. Multiple mechanisms underlie displacement of solitary Hawaiian Hymenoptera by an invasive social wasp. Ecology 91:3294-3302.

Wilson, E. O. 1971. The insect societies.

Wilson, E. O. 1976. Behavioral discretization and the number of castes in an ant species. 1:141-154.

Wilson, E. O. 1990. Success and dominance in ecosystems: the case of the social insects. Ecology Institute.

Wilson, E. O. 1996. Hawaii: A world without social insects 0:3-7.

Wilson, E. O. and B. Hölldobler. 2005. Eusociality: Origin and consequences. 102:13367-13371.

Woodard, S. H., B. J. Fischman, A. Venkat, M. E. Hudson, K. Varala, S. A. Cameron, A. G. Clark, and G. E. Robinson. 2011. Genes involved in convergent evolution of eusociality in bees. Proc. Natl. Acad. Sci. U. S. A. 108 : 7472-7477.

Wright, S. 1943. Isolation by distance. 28:114-138.

Wurm, Y., J. Wang, O. Riba-Grognuz, M. Corona, S. Nygaard, B. G. Hunt, K. K. Ingram, L. Falquet, M. Nipitwattanaphon, D. Gotzek, M. B. Dijkstra, J. Oettler, F. Comtesse, C. J. Shih, W. J. Wu, C. C. Yang, J. Thomas, E. Beaudoing, S. Pradervand, V. Flegel, E. D. Cook, R. Fabbretti, H. Stockinger, L. Long, W. G. Farmerie, J. Oakey, J. J. Boomsma, P. Pamilo, S. V. Yi, J. Heinze, M. A. D. Goodisman, L. Farinelli, K. Harshman, N. Hulo, L. Cerutti, I. Xenarios, D. Shoemaker, and L. Keller. 2011. The genome of the fire ant *Solenopsis invicta*. Proc. Natl. Acad. Sci. U. S. A. 108:5679-5684.

Wyman, M. J., A. D. Cutter, and L. Rowe. 2012. Gene duplication in the evolution of sexual dimorphism. Evolution 66:1556-1566.

Yanai, I., H. Benjamin, M. Shmoish, V. Chalifa-Caspi, M. Shklar, R. Ophir, A. Bar-Even, S. Horn-Saban, M. Safran, E. Domany, D. Lancet, and O. Shmueli. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21:650-659.

Yang, Z. H. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586-1591.

Zayed, A., N. L. Naeger, S. L. Rodriguez-Zas, and G. E. Robinson. 2012. Common and novel transcriptional routes to behavioral maturation in worker and male honey bees. Genes Brain Behav. 11:253-261.

Zdobnov, E. M., F. Tegenfeldt, D. Kuznetsov, R. M. Waterhouse, F. A. Simão, P. Ioannidis, M. Seppey, A. Loetscher, and E. V. Kriventseva. 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. 45:D744-D749.